Thesis

The ins and outs of black holes in AdS

VAN BREUKELEN, Rik

Abstract

This work starts by generalizing the traversable wormhole protocol from the thermofield double state black hole to time-shifted thermofield double state black holes and typical pure state black holes. The protocol is dependent on the time-shift and on the state respectively. The traversability of these states is evidence of the smoothness of the black hole horizons of these states. Inspired by these results, we show that the state-dependent construction of the interior of typical state black holes can be simplified using explicit time dependence. This reduces the state dependence to a single parameter.

Université de Genève
Departement de Physique Theorique

Faculté des sciences, Section de Physique
Professeur Marcos Mariño Beiras
Docteur Kyriakos Papadodimas

The ins and outs of black holes in AdS

THÈSE

présentée à la Faculté des Sciences de l'Université de Genève
pour obtenir le grade de
Docteur ès sciences, mention Physique

par

Rik van Breukelen

de

Zwolle (Pays-Bas)

Thèse N° 5488

Atelier d'impression de l'Université de Genève
Julliet 2020

UNIVERSITÉ DE GENÈVE

FACULTÉ DES SCIENCES

DOCTORAT ÈS SCIENCES, MENTION PHYSIQUE

Thèse de Monsieur Rik VAN BREUKELEN

intitulée :

«The ins and outs of black holes in AdS»

La Faculté des sciences, sur le préavis de Monsieur M. MARINO BEIRAS, professeur ordinaire et directeur de thèse (Section de mathématiques), Monsieur J. SONNER, professeur associé (Département de physique théorique) et Monsieur K. PAPADODIMAS, docteur (The Abdus Salam International Centre for Theoretical Physics, High Energy, Cosmology and Astroparticle Physics, Triest, Italy), autorise l'impression de la présente thèse, sans exprimer d'opinion sur les propositions qui y sont énoncées.

Genève, le 15 septembre 2020

Thèse - 5488 -

Le Doyen

N.B. - La thèse doit porter la déclaration précédente et remplir les conditions énumérées dans les "Informations relatives aux thèses de doctorat à l'Université de Genève".

Abstract

This work starts by generalizing the traversable wormhole protocol from the thermofield double state black hole to time-shifted thermofield double state black holes and typical pure state black holes. The protocol is dependent on the time-shift and on the state respectively. The traversability of these states is evidence of the smoothness of the black hole horizons of these states. Inspired by these results, we show that the state-dependent construction of the interior of typical state black holes can be simplified using explicit time dependence. This reduces the state dependence to a single parameter.

Acknowledgements

There are many people who have contributed in some way to the pages that you are now reading. Kyriakos Papadodimas has guided me from the beginning to the end. He helped me find balance between interesting and answerable problems, which lies on the knife's-edge between the trivial and unsolvable. It has been an adventurous journey we embarked on when we first met, and I will be forever thankful to him.

I would also like to thank my other co-authors Jan de Boer, Erik Verlinde, and Sagar Lokhande. Especially Sagar is deserving of my thanks. As a fellow PhD student, he allowed me to spar and make mistakes together, and to correct those mistakes in tedious calculations.

My gratitude also extends to Marcos Mariño for helping me overcome bureaucratic challenges, the SwissMap secretariat for helping me with everything government related, and the CERN theory group for hosting me.

Finally, I would like to thank my parents Sonja and Ebo, my brother Jelmer, and my partner Marieke for their unconditional love and support, and for, sometimes, even trying to understand what I was working on.

Résumé en français

Cette thèse examine si l'horizon d'un trou noir est lisse. Nous le faisons dans le cadre simplifié de la correspondance AdS / CFT, où nous pouvons traduire les questions relatives à la géométrie dans AdS en un problème de la CFT.

Le premier outil que nous utilisons est le *traversable wormhole protocol*. Ce protocole a été décrit pour la première fois pour *thermofield double state*, l'état idéalisé d'un trou noir. Nous généralisons le protocole pour l'appliquer à une classe d'états appelés *time-shifted thermofield double states*. Nous montrons que cette classe d'états est aussi fluide que le *thermofield double state* en envoyant un signal d'un côté du trou de ver et en le recevant de l'autre côté.

Ensuite, nous examinons les trous noirs typiques à l'état pur. Les États présentent un grand intérêt car les *firewall paradoxes* interdisent un horizon lisse décrit par des opérateurs indépendants des États. Ces paradoxes peuvent être évités en faisant appel à des opérateurs dépendants de l'état, également appelés *mirror operators*.

Nous combinons les *mirror operators* avec le *traversable wormhole protocol* pour généraliser le protocole à des états typiques. Nous obtenons la preuve de la finesse de l'horizon en envoyant un signal de l'intérieur du trou noir et en le recevant à l'extérieur. Nous montrons également comment extraire des informations d'un trou noir en envoyant un signal dans le trou noir et en le récupérant ultérieurement.

Nous concluons également que la géométrie d'état typique inclut une partie de la région située au-delà de l'intérieur de la géométrie étendue AdS-Schwarzschild. Cette région est décrite par les *mirror operators*. L'isométrie de Killing se déplace en fonction de l'évolution temporelle dans cette région. Cela signifie que les *mirror operators* doivent avoir une dépendance temporelle explicite pour éviter les courbes temporelles fermées.

Dans la dernière partie, nous montrons que la dépendance temporelle explicite est suffisante pour éviter les *firewall paradoxes*. Cela nous permet de réduire la dépendance d'état des *mirror operators*. à un paramètre unique. Nous donnons une construction potentielle à ces nouveaux *mirror operators*. en utilisant le cône naturel décrit par la théorie de Tomita-Takesaki.

Contents

Chapter 1

Introduction

A black hole is an object so heavy that even light cannot escape from it. These objects where considered well before the appearance of modern physics. Einstein's general relativity and the solution found by Schwarzschild gave these ideas a modern interpretation. With this mathematical description, it became obvious that there is point of no return around a black hole. You would have to go faster than the speed of light to escape beyond that point. This surface is called the Schwarzschild radius for that particular solution of the Einstein equations. It was initially thought that this surface was problematic as some term in Einstein equations became infinite at that point. However, it was realized after few years that was a coordinate singularity, which is avoidable by the right coordinate transformation.

Another couple of years later and the problems surrounding the point of no return, the black hole horizon, reappear. It was proven that black holes can be fully characterized by mass, spin, and charge. This suggests that information is lost when thrown into a black hole. It was Hawking who showed that black hole have entropy,

$$S = \frac{A}{4G}\frac{k_B c^3}{\hbar}. \tag{1.1}$$

Hawkings derivation showed that black holes have entropy, so they store information, and that they have a temperature, so they radiate. However, the resulting radiation is perfectly thermal and information disappears. This information loss is not allowed by quantum mechanics and years of discussion followed.

That the black hole entropy is proportional to the area was the first hint of the holographic principle. This principle says that you should be able to describe the interior of a system by the degrees of freedom at the boundary. The AdS/CFT correspondence does this for AdS spaces and tells us that information is conserved. However, many questions remain about how information is conserved and retrieved even in these simplified case. For example, what do you experience when you fall into the black hole and cross the horizon?

We need to translate the question whether the black hole horizon is smooth from AdS to the CFT by making use of the tools of the AdS/CFT correspondence. The traversable wormhole protocol [5] provides a thought experiment that can be expressed as CFT correlators, while probing the interior of a black hole. We generalize this protocol to apply also to apply to time-shifted thermofield double states in chapter 2, which coincides with the discussion in:

> [1] van Breukelen, Rik and Papadodimas, Kyriakos, *Quantum teleportation through time-shifted AdS wormholes*, JHEP **08** (2017) 142 [Arxiv: 1708.09370]

The generalization of the protocol needs to depend on the time-shift in order to work properly. It is, therefore, state dependent in this simple form.

In chapter 3 we generalize the traversable wormhole protocol further and apply it to typical pure state black holes. We made use of state-dependent interior operators [6, 7, 8] to achieve this in the following papers:

> [2] De Boer, Jan and van Breukelen, Rik and Lokhande, Sagar F. and Papadodimas, Kyriakos and Verlinde, Erik, *On the interior geometry of a typical black hole microstate*, JHEP **05** (2019) 010 [Arxiv: 1804.10580]

and

> [3] De Boer, Jan and van Breukelen, Rik and Lokhande, Sagar F. and Papadodimas, Kyriakos and Verlinde, Erik, *Probing typical black hole microstate*, JHEP **01** (2020) 062 [Arxiv: 1901.08527]

The state dependent operators describe the interior of a black hole. Moreover, they describe a part of the geometry beyond the interior of a black hole. An interesting consequence of this is that the interior operators must be explicitly dependent on the time. This explicit time dependence is needed to avoid closed timelike curves.

We direct the reader interested in the SYK model to section 3.3, where we apply the traversable wormhole protocol to the SYK model.

In chapter 4, we exploit the explicit time dependence of the interior operators to reduce the state-dependence of the interior operators to a single parameter, which was first described in:

> [4] van Breukelen, Rik, *Black Hole State Dependence as a Single Parameter*, JHEP **04** (2020) 210 [Arxiv: 1910.00036]

This single parameter state-dependence increases the naturalness of the construction and is similar to the state-dependence in the case of the time-shifted thermofield double states.

Each chapter contains a short review of the needed background material. Therefore, all chapters can be read independently.

Chapter 2

Teleportation through time-shifted wormholes

It has been demonstrated that the Einstein-Rosen bridge of the eternal AdS black hole can be made traversable via a particular double-trace deformation of the boundary CFTs [5]. Before the deformation, the two boundary theories were non-interacting and placed in a specific entangled state $|\Psi_{\text{tfd}}\rangle = \sum_E \frac{e^{-\frac{\beta E}{2}}}{\sqrt{Z}}|E,E\rangle$. The deformation creates shockwaves in the bulk, with negative average null energy, which shrink the horizon of the black hole a little, allowing a particle to traverse the wormhole from one asymptotic region to the other. The deformation can also be formulated as a quantum teleportation protocol between the two CFTs [5, 9]. This setup has provided evidence for the smoothness of the horizon of the eternal black hole and for the ER=EPR proposal [10].

In this chapter, we consider a similar experiment on a large class of states with different details in the entanglement between the CFTs. These states are of the form $|\Psi_T\rangle = e^{iH_R T}|\Psi_{\text{tfd}}\rangle$, where T is a parameter controlling the entanglement. While these states are as entangled as $|\Psi_{\text{tfd}}\rangle$, they are *different* quantum states. We argue that the double-trace deformation (and the quantum teleportation protocol) can be modified to apply to each one of the states from this class. This provides evidence that they all have a smooth horizon.

This simple observation has some interesting implications. First, states of this family with $T > 0$ can in principle be used in a lab setup to allow an observer crossing the wormhole to travel far in the future in finite amount of proper time. During the trip the observer is mostly in free fall. Second, for states with $T < 0$, the bulk observer experiences evolution by finite proper time, while the elapsed time in the lab can become very small. Finally, the fact that we can establish the smoothness of this class of time-shifted states is of some interest for the firewall paradox [11, 12, 13] and the state-dependent proposal of [6, 7, 8, 14, 15, 16].

We emphasize that the CFT correlators needed to support the claims of this chapter are isomorphic to those relevant for [5, 9]. Hence, proving the traversability of the wormhole in the time-shifted states is equivalent to the same proof for the TFD.

This chapter is organized as follows: in section 2.1, we discuss the basic setup used by [5, 9], which is the basis of the traversable wormhole. In section 2.2, we argue that time-shifted wormholes can be made traversable in a similar manner. In sections 2.3 and 2.4 we discuss this setup from a laboratory and a quantum-teleportation point of view. Finally,

in section 2.5 we discuss some connections to the firewall paradox and state-dependence.

2.1 Traversable AdS wormholes

In [17] it was proposed that two non-interacting copies of the same holographic CFT placed in the "thermofield" (TFD) entangled state

$$|\Psi_{\text{tfd}}\rangle = \sum_E \frac{e^{-\frac{\beta E}{2}}}{\sqrt{Z}} |E\rangle \otimes |E\rangle$$

are dual to the eternal black hole in AdS. This gravitational background can also be thought of as a wormhole connecting two asymptotic AdS regions. However, in this setup the wormhole is non-traversable, which is important for consistency given that the boundary CFTs are non-interacting and hence no information can be exchanged between them.

It was realized in [5] that the wormhole can become traversable by coupling the two CFTs with a double-trace perturbation $e^{ig\mathcal{O}_L\mathcal{O}_R}$, which is turned on for a short time around $t = 0$. Here $\mathcal{O}_{L/R}$ is a simple operator in the two corresponding CFTs. By selecting the sign of g appropriately, the perturbation creates negative null energy shockwaves falling into the black hole from both sides, see figure 2.1. This shrinks the horizon a little. As a result, an observer who dives from the left CFT at $t = t_{\text{in}} < 0$ towards the black hole emerges on the right side and reaches close to the right boundary at $t = t_{\text{out}} > 0$. Here both $|t_{\text{in}}|$ and t_{out} are of the order of the scrambling time $\beta \log S$. The details may depend on the theory and the form of the shockwave.

This setup is interesting because it allows us to probe the space-time in the interior of the wormhole purely in terms of two-sided correlators of standard CFT operators. Directly probing the black hole interior from the CFT is more difficult, because we first have to define approximately local operators in AdS, which is non-trivial, especially behind the horizon. The setup of [5] bypasses the need to define these local bulk operators, as it probes the interior indirectly. The observer is created by a local CFT operator $\phi_L(t_{\text{in}})$, the perturbation is generated by CFT operators $\mathcal{O}_L(0)\mathcal{O}_R(0)$ and the outgoing observer is detected by a local CFT operator $\phi_R(t_{\text{out}})$. So the question of what happens to an observer falling through this wormhole can be translated into a computation of correlators of local CFT operators, the analogue of an S-matrix element in AdS. These well-defined (though difficult to compute in practice[1]) CFT correlators can in principle provide evidence for the smoothness of the horizon of the eternal black hole and of the proposal [17].

In [5] it was pointed out that this protocol is related to quantum-teleportation, and in [9] this quantum-teleportation protocol was made more explicit. First the observer is placed in the left CFT at $t = t_{\text{in}}$. At $t = 0$ the operator $\mathcal{O}_L$ is measured. Depending on the resulting eigenvalue o_L, the unitary $e^{igo_L\mathcal{O}_R}$ is applied on the right CFT. Then the quantum state of the system at time $t = t_{\text{out}}$ contains the observer emerging from the black hole into the right asymptotic AdS region. See also [18].

2.2 Time-shifted wormholes

In this chapter, we will work under the assumption that the TFD state can be made traversable for a semi-classical observer, as argued in [5, 9]. Using this as our starting

[1] In [9] relevant correlators were computed in the SYK model.

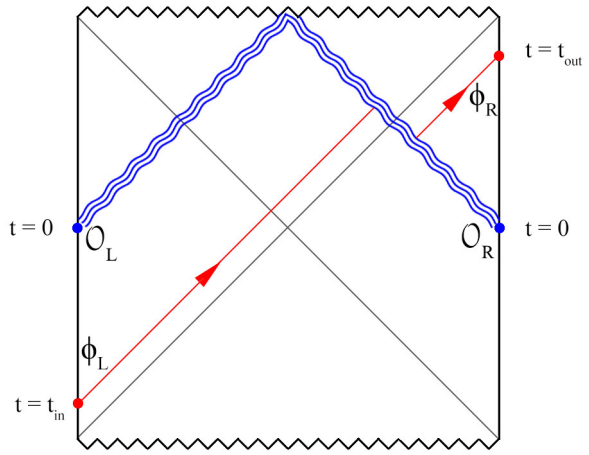


Figure 2.1: A double trace deformation creates a shockwave which displaces the probe ϕ, allowing it to escape from the black hole. The coordinates are discontinues at the shockwave, while the path of the probe is smooth.

postulate, we point out that there is a large class of other states with similar behaviour. These are states of the form

$$|\Psi_T\rangle \equiv e^{iH_R T}|\Psi_{\text{tfd}}\rangle = \sum_E \frac{e^{-\frac{\beta E}{2}}}{\sqrt{Z}} e^{iET}|E\rangle \otimes |E\rangle \tag{2.1}$$

It is important to realize that these are different quantum states from $|\Psi_{\text{tfd}}\rangle$, due to the energy-dependent phases.

The bulk interpretation of these time-shifted states, is that they are related to the usual eternal black hole by a large diffeomorphism, see for example [15]. This is a diffeomorphism which acts as a time translation on the right boundary, but trivially on the left boundary. Since this is a large diffeomorphism (allowed by the boundary conditions), we are not supposed to mod-out by it. Instead, it maps a physical state to a different physical state. The states (2.1) can be represented as the usual eternal AdS black hole, but where the wormhole is "anchored" at different points in time on the two boundaries.

i) T>0

We argue that for every choice of $T > 0$ the traversable wormhole protocol of [5] can be implemented: at $t = t_{\text{in}} < 0$ an observer jumps into the left CFT. At $t = 0$ we briefly couple the two CFTs by the operator $e^{ig\mathcal{O}_L(0)X_R(0)}$, where now $X_R(0) = e^{iH_R T}\mathcal{O}_R e^{-iH_R T}$ is the "precursor" of the operator $\mathcal{O}_R(T)$. Then at time $t = T + t_{\text{out}}$ the observer will come out in the right CFT, in exactly the same form as in the original experiment on the state $|\Psi_{\text{tfd}}\rangle$. See figure 2.2. Alternatively, we could have used a precursor on the left, i.e. coupling the two CFTs with $e^{igY_L(T)\mathcal{O}_R(T)}$ where $Y_L(0) = e^{-iH_L T}\mathcal{O}_L e^{iH_L T}$, or some combination of left and right precursors at time t satisfying $t_{\text{in}} < t < T$. The details of how the result of this experiment is isomorphic to that of the TFD is explained in appendix A.

We emphasize that this statement is *exact*, even if T is appreciably large. In other words, provided we accept that the protocol leads to a smooth traversable wormhole for the observer falling into the TFD, then same can happen for all the other states, without

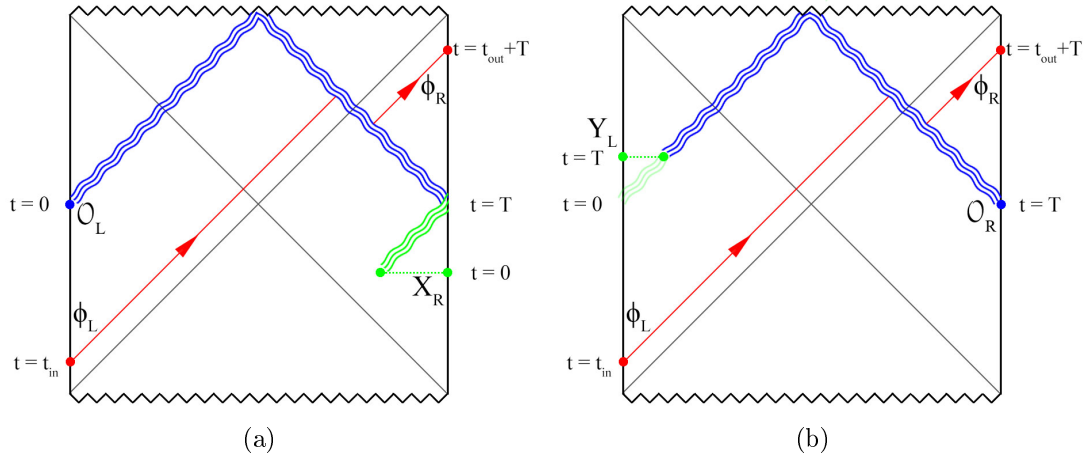


Figure 2.2: a) In the time-shifted wormhole, with $T > 0$, we need to act with a more complicated operator X_R to receive the probe. b) Similar results can be achieved by using a precursor on the left CFT. Note that the Penrose diagrams can be misleading for precursors, because they may have a more involved bulk interpretation, see for example [19]. However, the quantum state on the boundary after the end of the experiment can be reliably predicted.

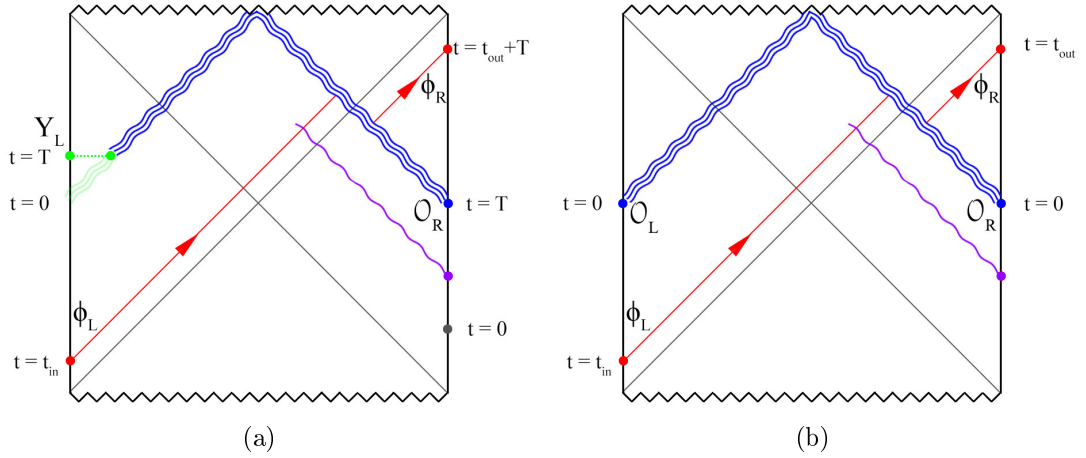


Figure 2.3: a) The memory of the probe can be modified by sending an early perturbation from the right. b) The same setup for the thermofield state.

any approximation. By tuning T we can arrange that the observer will emerge significantly later in the future. Moreover, the quantum state of the observer after emerging in the right CFT will be exactly the same — simply displaced in time. In particular her memories, and the amount of proper time that she will think has elapsed, will be the same[2] and independent of T.

While we can show that the observer emerges in the right CFT with memories of a smooth crossing of the wormhole, there is a logical possibility that the following scenario took place: during the crossing the observer actually experienced some unpleasant parts — for example a firewall — which killed her upon impact. Then the dynamics of the system

[2] If we consider a big black hole in AdS whose radius $R_{\text{bh}} \sim R_{\text{AdS}}$ then the elapsed proper time will also be of order R_{AdS}.

"resurrected" the observer on the right side and in a state with memories corresponding to a smooth crossing. This scenario may sound un-natural, but in some sense it is not so difficult to realize mathematically: for instance imagine an observer living inside a quantum system and that we act at $t = t_0$ with a unitary U which kills him. At a later time t_1 we act with the precursor $e^{-iH(t_1-t_0)}U^{-1}e^{iH(t_1-t_0)}$. Then the quantum state of the system for $t > t_1$ is the same as what it would have been, had we not acted with the first unitary which killed the observer. In this sense the sequence of U at t_0 and its inverse precursor at $t = t_1$ kills and resurrects the observer. Moreover, when resurrected the observer has no memories of the fact that he had been killed. Notice that the unitary U and its inverse precursor do not have to be fine-tuned with respect to the initial state of the observer in order to be able to resurrect him.

In our setup the meaning of this question is whether two-sided CFT correlators can exclude the possibility that the observer was killed when falling into the black hole from the left and resurrected when emerging in the right CFT.

In order to directly address this question we would have to define local bulk observables which would be able to tell us what really happened in the middle of the bulk spacetime. As an easier alternative, we can send early signals from the right to probe the path of the observer, see figure 2.3. These signals must be sufficiently weak, to avoid killing the observer or pushing the observer [20, 21] outside the window in which the coupling between the CFTs allows the extraction of the observer. We can then study if these signals modify the final quantum state of the outgoing observer in the way which is expected from effective field theory. If they do so, then we get additional evidence, though not definitive proof, that nothing dramatic happened to the observer while crossing. Then this becomes again a statement of CFT correlators, which could in principle be computed.

We can see that these CFT correlators in the time-shifted TFD states are again isomorphic to the same correlators in the TFD state, provided that the signals from the right are sent with the appropriate time-shift. Hence if nothing strange happens to an observer crossing the TFD state as claimed by [5, 9], then the same will be true for the time-shifted states. See appendix A for some details.

ii) $\mathbf{T<0}$

We can also consider states with $T < 0$, with $|T| < t_{\text{out}}$. We can couple the two CFTs with the operator $e^{igO_L(0)X_R(0)}$, where again X_R is the precursor of the operator $\mathcal{O}_R(T)$. Provided that $|T| < t_{\text{out}}$, the observer will emerge in the right CFT at $t = T + t_{\text{out}} > 0$. Notice that in this setup the total lab time it takes for the observer to cross the wormhole is $T+|t_{\text{in}}|+t_{\text{out}}$, which is less than $|t_{\text{in}}|+t_{\text{out}}$. So the crossing of the observer is accelerated for the lab frame, even though the proper time according to the observer is exactly the same.

Actually, we can shorten the lab time even more as follows. We throw the observer from the left at $t = t_{\text{in}}$ and we couple the two CFTs at $t = t_{\text{in}}+\epsilon$ by the operator $e^{igY_LX_R}$, where both are precursors. Then the observer comes out at $t = T+t_{\text{out}}$. Causality requires that $T+t_{\text{out}} > t_{\text{in}}+\epsilon$. This means we can push T towards the negative values all the way to $T_{\text{min}} = -|t_{\text{in}}| - t_{\text{out}} + \epsilon$. In this case the observer emerges on the right CFT almost immediately, even though according to her own experience the same (finite) amount of proper time as before has elapsed.

Notice that the full bulk interpretation of these protocols may be complicated due to

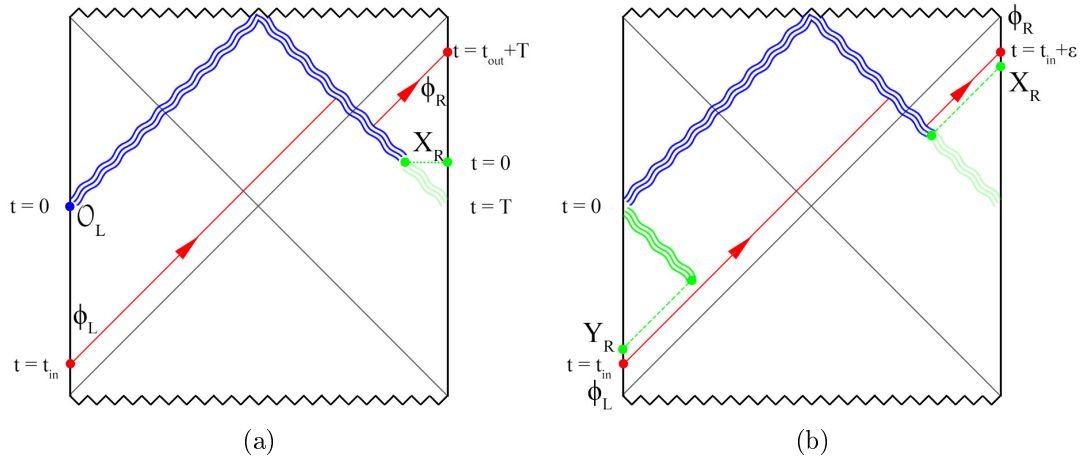


Figure 2.4: a) In the time-shifted wormhole, with $T < 0$, we can still recover the probe provided that T is not to large. b) The extreme case in which we receive the probe almost immediately.

the use of the precursors, which are complicated non-local operators. On the other hand we can reliably predict the exact quantum state of the observer — with memories of smooth crossing and finite elapsed proper time — as she emerges on the right CFT.

2.3 Laboratory point of view

In order to avoid possible confusions regarding the meaning of the time-shift in (2.1), it is useful to think of the experiment in the following way. We imagine a laboratory in a universe where gravity does not play an important role. We have two identical CFTs realized in some material in the laboratory. These are supposed to be holographic CFTs dual to gravity in AdS. There is only one common time, the laboratory time t. Each of the CFTs evolve with its own Hamiltonian, but since the CFTs live in the laboratory we identify the CFT time with the lab time $t_L = t_R = t$. The CFTs are prepared to be in a specific entangled state (2.1). Hence it is more appropriate to think of the parameter T as a "dial" that selects the initial state, rather than a time-shift. Of course preparing two CFTs in the TFD state or in one of its time-shifted cousins would be very difficult in practice, but possible in principle.

The laboratory technician can prepare a protocol where the observer is first injected into the left CFT at $t = t_{\text{in}}$. At $t = 0$ the lab technician couples the two CFTs by the operator mentioned previously. Then at $t = T + t_{\text{out}}$ the observer emerges in the right CFT. From the point of view of the observer only a finite proper time elapses which is independent of T, but from the point of view of the lab the elapsed time is $T + |t_{\text{in}}| + t_{\text{out}}$, where T can be arbitrarily large. Moreover throughout this experiment the observer is in free-fall, except for the (mild) interaction with the shockwave. We emphasize that the subjective experience of the observer is independent of the value of T and in particular the strengh of the interaction with the schockwave is also independent of T.

It is interesting to notice that from the boundary point of view, the quantum information of the observer jumps from the left to the right CFT at $t = 0$ when we couple the two CFTs. Then it stays scrambled in the right CFT for a long time, until it emerges in

simple form at $t = T + t_{\text{out}}$. For instance, suppose that the observer on the left CFT carries a spin which is maximally entangled with some external reference spin. For $t < 0$ the purification of the reference spin is in the left CFT. Right after $t > 0$ the purification is in the right CFT but in scrambled form. Eventually at $T + t_{\text{out}}$ the purification of the reference spin is in the right CFT in terms of a simple spin carried by the observer.

2.4 Quantum teleportation

The double-trace perturbation introduced in [5] can be slightly modified to be interpreted as a quantum teleportation protocol. In [9] this was described as follows: we make use of the fact that anything we do on the left boundary after acting with the double trace perturbation cannot affect the right boundary. For example, we could measure $\mathcal{O}_L$ just after the perturbation. Because $\mathcal{O}_L$ and $e^{ig\mathcal{O}_L\mathcal{O}_R}$ commute it would be equivalent to measure $\mathcal{O}_L$ just before the perturbation and then perturb by $e^{io_L\mathcal{O}_R}$, where o_L is the eigenvalue measured. Therefore, instead of acting with the double trace perturbation, the lab technician can implement the following protocol. First he releases the probe at $t = t_{\text{in}}$ in the left CFT. Then he measures $\mathcal{O}_L$ at $t = 0$ and project onto one if its eigenstates with resulting eigenvalue o_L. Then he acts with a unitary $e^{igo_L\mathcal{O}_R}$ at $t = 0$ on the right CFT. The right CFT density matrix at the end of the teleportation protocol will be the same as the one in the double trace protocol, while the one on the left boundary will be different. Notice that in the step of recording o_L and selecting accordingly the unitary on the right we have the transfer of *classical* information from left to right, which is a part of a quantum teleportation protocol.

The quantum teleportation protocol can be immediately realized for the time-shifted states: the lab technician first measures $\mathcal{O}_L$ at time $t = 0$. Using the resulting eigenvalue o_L he applies at $t = 0$ the unitary $U = e^{igo_L X_R}$ on the right CFT. Here $X_R(t = 0)$ is the complicated precursor corresponding to the simple operator $\mathcal{O}_R(t = T)$. Finally at time $t = T + t_{\text{out}}$ the density matrix of the right CFT will be the same as in the experiment on the TFD at time $t = t_{\text{out}}$. This protocol is possible in principle, but it requires the use of the complicated operator X_R.

In the case that $T > 0$ the lab technician can avoid having to use a complicated precursor, by performing an alternative "time-delayed quantum teleportation protocol". He releases the probe in the left CFT at $t = t_{\text{in}}$ ii) then he projects onto an eigenstate of $\mathcal{O}_L$ at $t = 0$, recording the eigenvalue o_L. Then he waits until $t = T$ and he acts with a simple unitary $U = e^{igo_L\mathcal{O}_R}$ at $t = T$. Finally he considers the right CFT density matrix at $t = t_{\text{out}} + T$. This protocol has the advantage that we do not have to use complicated precursors. We notice that we cannot use this protocol when $T < 0$, as we would need to apply a unitary before the measurement of $\mathcal{O}_L$.

2.5 Comments on state-dependence and the firewall

The firewall paradox can be understood in its most precise formulation in terms of *typical* pure states of a one-sided black hole in AdS. The argument starts by assuming that typical pure states have a smooth interior. It is then assumed that there should exist some fixed linear operators acting on the Hilbert space, which correspond to local semiclassical observables behind the horizon. It is then shown that, according to bulk effective field

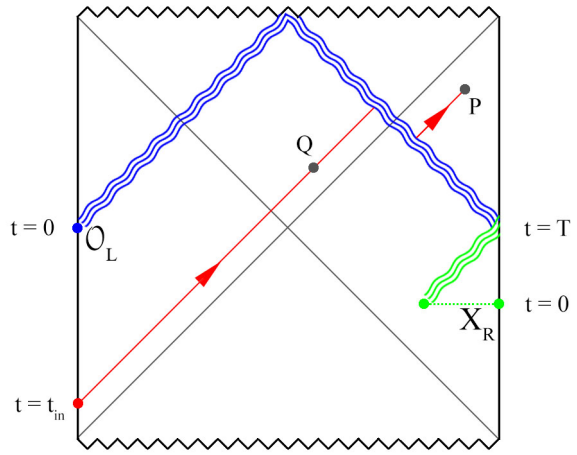


Figure 2.5: Local operators at points P, Q are state-dependent

theory, these observables would have to obey an algebra which is inconsistent with the density of states in the CFT [12, 13]. For this argument to work, it is important that we demand a smooth interior for a large class of states, i.e. for typical states. If we only look at a small number of states, then the paradox becomes less sharp. In [6, 7, 8, 14, 15, 16] it was proposed that the paradox in its strongest form, i.e. for typical states, can be resolved by allowing the interior operators to depend on the state.

The smoothness of TFD state, as demonstrated by [5, 9] does not disprove the firewall, as the TFD is one particular state, while the firewall paradox becomes relevant when we consider *many* states. However, in [15] it was pointed out that a version of the firewall paradox can also be formulated if we consider the entire family of time-shifted TFD states $e^{iH_R T}|\Psi_{\text{tfd}}\rangle$ for all $T \in \mathbb{R}$. It was shown in [15] that if we demand smoothness for all of these states, then we run into a firewall-like paradox, unless we accept that the interior operators are state-dependent. The argument of [15] was based on the assumption of smoothness for all time-shifted TFD states. This seems very plausible from the bulk point of view, since they are all related to TFD by a large diffeomorphism. However, it would be more satisfying if there was more direct evidence for the smoothness of the time-shifted TFD states.

In this chapter, we argued that by applying the teleportation protocol [5, 9] to the time-shifted states for all $T \in \mathbb{R}^+$, we find that all of them have a smooth interior. This disproves the firewall within a class of states where one would naively expect some firewall-like behavior[3]. A natural explanation is that the interior operators in these states are indeed state-dependent.

The class of time-shifted TFD states, together with the perturbation which allows the particle to escape the horizon, raise an interesting aspect of state-dependence for observables *outside the horizon*. Consider a local bulk operator at point P in figure 2.5. According to the infalling observer, this point is reached by diving in from the left CFT at t_{in} and freely-falling for a fixed amount of proper-time. For the infaller this relational prescription of the point P is the same for all states, independent of T. However, the measurement of the operator at P takes place at laboratory time $t = T + t_{\text{out}}$. So this

[3]The argument of [15] leads to a firewall-like paradox, even if we restrict to the family of states with $T > 0$. To formulate this paradox we need to be able to take T up to a time scale of order e^S.

local operator at P can be represented as *the same* operator in the Schrödinger picture, however — depending on the microstate — it is applied by the infalling observer on the Schrödinger-picture Hilbert space corresponding to a different time.

We notice that the same property holds for local operators inside the horizon for this class of states, for example for a local operator at point Q. It is interesting to understand how this happens from the point of view of the infalling semiclassical observer, i.e. how does she naturally identify the correct moment in time where the operators have to be applied.

Discussion

We investigated an extension of the traversable-wormhole protocol of [5, 9], which has interesting physical interpretations. We argued that using a larger class of entangled states, the time-shifted thermofield states, can lead to experiments involving time-travel in the lab.

General relativity allows time-travel to the future, by hovering near the horizon for a while and then flying away an observer can travel to the future. However, in order to move far in time, this method is not very pleasant, as it requires large proper accelerations. In this chapter, we described a more comfortable time-machine based on quantum entanglement. From the point of view of the observer the experience is pleasant, even if the desired time-difference is large.

We notice that when the time shift T becomes of the order of, or larger that the Poincare recurrence time, then the physical interpretation of the process must be done more carefully, since the observer may come out earlier than $T + t_{\text{out}}$, in the "previous Poincare recurrence".

These long timescales are important when considering TFD states with random phases, $|\Psi_\theta\rangle = Z^{-1/2} \sum e^{-\frac{\beta E}{2}} e^{i\theta_E} |E\rangle \otimes |E\rangle$, where the θ_E's are random phases. These states can be well approximated by a timeshifted state, where the timeshift is determined by matching the phases of the state. In this approximation we may need to exclude a small number of states, as energy eigenstates with rational relations may never align their phases under a timeshift. The amount of timeshift can be estimated by Kac's lemma [22], which shows that these timeshifts are of the order of the Poincare recurrence timescale $T = O\left(e^{e^S}\right)$.

We also argued that for certain states with $T < 0$ we can retrieve the observer almost immediately. One might worry that we can create a very fast computer by sending a computer through the wormhole, while there are fundamental bounds on computation speeds [23]. However, the CFTs creating the wormhole should be included as part of the computer, which will presumably respect the bounds.

It would be interesting to investigate the traversable wormhole protocol for more general entangled states of two CFTs. One particular class of such states would be superpositions of time-shifted thermofield states. Finally it would be interesting to investigate the possibility of traversing a single-sided black hole. In the case of the SYK model this was discussed in [24]. Another class of candidate states in general holographic CFTs, which could be used as a starting point was proposed in [25].

Chapter 3

The interior geometry of a typical black hole

The black hole information paradox is a long-standing open problem, which is related to the smoothness of the black hole horizon [26, 11]. The AdS/CFT correspondence provides an ideal setting to investigate the issue of smoothness. Large typical black holes in AdS are expected to be dual to typical high-energy pure states in the dual CFT. These typical black holes are approximately in equilibrium and hence do not evaporate. Even then, it is challenging to reconcile the smoothness of the horizon with unitarity of the dual CFT [12, 13, 27]. In this chapter, we make some inroads into investigating the geometry of such a typical black hole microstate.

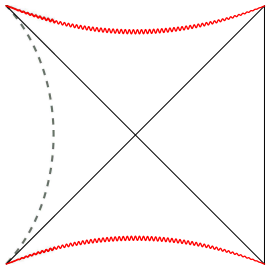


Figure 3.1: A proposal for the geometry dual to a typical black hole microstate.

Owing to robust arguments in the AdS/CFT framework, it is widely believed that at large N the geometry of a typical black hole microstate contains at least the region exterior to the black hole horizon, which is described by the AdS-Schwarzchild metric. The question then is: do there exist any other regions in the geometry dual to a typical black hole microstate? It seems reasonable that any proposed answer to this question needs to satisfy two constraints: (1) the geometry in the exterior should be that of the AdS-Schwarzschild black hole, (2) the geometry should manifest the approximate time-translation-invariance of the typical pure state in the CFT, through the existence of an approximate timelike Killing isometry. We discuss the time-translation-invariance of typical pure states in Section 3.1.1.

In [11, 12, 13], it was suggested that the geometry of a typical black hole microstate contains only the exterior region, which gets terminated at the horizon by a firewall.

However, for large typical black holes, the curvature near the horizon is low. Thus, this proposed solution demands a dramatic modification of general relativity and effective field theory in regions of low curvature.

In this chapter, we will explore the possibility that the bulk geometry of a typical pure AdS black hole microstate contains part of the extended AdS-Schwarzchild diagram, as shown in figure 3.1. Since the dual of this geometry is a typical pure state in a single CFT, the Penrose diagram cannot be extended arbitrarily to the left and there is no "left" CFT. The dotted line in figure 3.1 denotes a surface beyond which the geometry is not operationally meaningful. We will discuss the interpretation and other features of this geometry in later sections.

The proposal that the geometry dual to a typical microstate includes parts of the black hole, white hole and left regions, as depicted in figure 3.1, is suggested by the existence of CFT operators which have the right properties to represent these regions. These are the *mirror operators*, denoted by $\widetilde{\mathcal{O}}$, a set of *state-dependent* operators identified by an analogue of the Tomita-Takesaki construction applied to the algebra of single-trace operators $\mathcal{O}$ in the CFT. At large N, the operators $\widetilde{\mathcal{O}}$ commute with usual single trace operators and they are entangled with them. They are the natural candidates to describe the left region of the extended black hole geometry. The black hole interior and white hole region would then be reconstructed by a combination of $\mathcal{O}$ and $\widetilde{\mathcal{O}}$.

Naively, the left region would be inaccessible from the CFT, at the level of effective field theory. However, starting with the work of Gao, Jafferis and Wall [5] and further work [9, 24], a new approach has been identified for probing the space-time beyond the horizon, including the left region. This new approach, which was formulated in the framework of the two-sided eternal black hole, is based on the observation of [5] that in the case of the two-sided eternal black hole there are perturbations of the boundary CFTs of the form $\delta H = \mathcal{O}_L \mathcal{O}_R$, which can create negative energy shockwaves which can violate the average null energy condition and allow particles to traverse the horizon. This effect is related to the quantum chaotic behavior of out-of-time-ordered correlators (OTOC) at scrambling time in the boundary CFT [28, 29].

In this chapter, we provide evidence for the conjectured geometry of figure 3.1 for the one-sided black hole, by perturbing the CFT Hamiltonian by the state-dependent operators $\widetilde{\mathcal{O}}$, in the schematic form $\delta H = \mathcal{O}\widetilde{\mathcal{O}}$. These perturbations allow particles that are localized in the left region of the geometry dual to a pure microstate, to traverse the black hole region and emerge in the right region and get directly detected by single-trace CFT operators. This is schematically shown in figure 3.2. We emphasize that the use of state-dependent operators from the point of view of the boundary CFT falls within the standard framework of quantum mechanics[1] and is logically independent from the question of how the infalling observer can use these operators.

Using these state-dependent perturbations by mirror operators, we argue that the consistency of the space-time geometry proposed in this chapter, and shown in figure 3.1, requires as a *necessary* condition that CFT correlators of *ordinary* CFT operators should obey the following property: *the effects of quantum chaos, which become important in out of time order correlators (OTOC) at scrambling time, should be the same — to leading*

[1] We can imagine that the boundary observer has prepared many identical systems. By performing measurements in many of these copies he can determine the exact microstate. Then, the observer can prepare an experimental device that acts with the operators $\widetilde{\mathcal{O}}$ relevant for that microstate and apply them to one of the identically prepared copies which has not been previously measured.

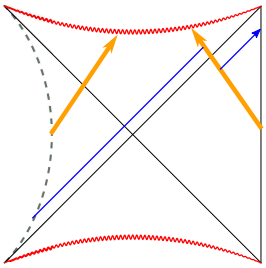


Figure 3.2: Perturbations of H_{CFT} by state-dependent operators can create negative energy shockwaves (yellow), which allow a probe from the left region (blue) to get detected by simple operators in the CFT.

order at large N — in typical pure states as in the thermal ensemble. We conjecture that this is true in large N holographic CFTs and provide some indirect evidence. Notice that this conjecture about the OTOC in pure states is a statement which is independent of the bulk interpretation, and in principle it can be either verified or falsified by CFT methods. A verification of this CFT conjecture would be a necessary condition for the validity of the geometry of figure 3.1.

Finally, we argue that the mirror operators $\widetilde{\mathcal{O}}$ can be used to implement an analogue of the Hayden-Preskill protocol [30], in its formulation given in [9]. Information thrown into an AdS black hole, which was originally in a typical state, can be recovered by deforming the CFT Hamiltonian by $\mathcal{O}\widetilde{\mathcal{O}}$ and then measuring a mirror operator after scrambling time. An analogue of this protocol can be applied to black holes in flat space after Page time. Then the mirror operators are mostly supported on the early Hawking radiation, which forms the larger fraction of the total Hilbert space. Interestingly, the protocol then becomes an analogue of the Hayden-Preskill protocol. The complicated nature and state-dependence of operators $\widetilde{\mathcal{O}}$ is consistent with the fact that for the application of the Hayden-Preskill decoding protocol, the observer must have knowledge of the initial black hole microstate and apply a state-dependent decoding procedure.

The plan of the chapter is as follows: in section 3.1 we provide details about the conjectured geometry of a typical black hole microstate. In section 3.2 we describe how time-dependent perturbations of the CFT Hamiltonian using state-dependent operators allows us to probe the interior. In section 3.3 we formulate and investigate some of our general statements in the SYK model. In section 3.4 we discuss the technical conjecture about the chaotic OTOC correlators in pure states and provide some evidence for its validity. In section 3.5 we discuss the connection of our experiments with the Hayden-Preskill protocol.

Part of the results of this chapter were reported in a shorter note [2]. Other recent works which investigate the region behind the horizon of special, *atypical* pure states include [31, 32, 33].

3.1 On the interior geometry of a typical state

In this section we present a conjecture for the bulk dual of a typical black hole microstate in the framework of AdS/CFT correspondence. We review the construction of *mirror*

operators, CFT operators that may describe the region behind the horizon. We also discuss how time-dependent perturbations of the CFT Hamiltonian by mirror operators can create excitations behind the horizon.

3.1.1 Typical Black Hole Microstates and the "Mirror Region"

The Penrose diagram of an AdS black hole formed by collapse is shown in figure 3.3. In this

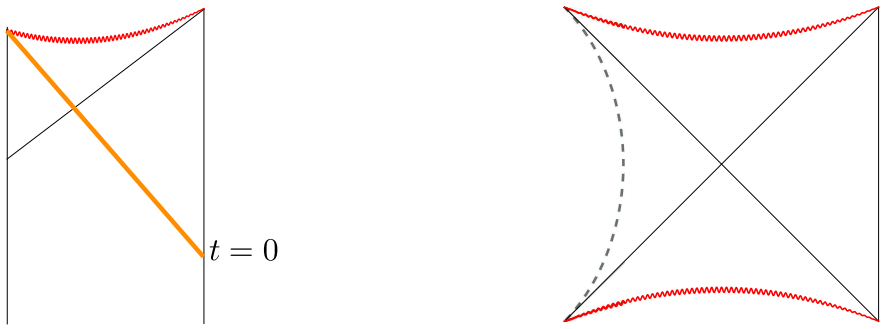


Figure 3.3: Left: AdS black hole formed by collapse. Right: conjectured bulk geometry of a typical black hole microstate in a single CFT. The left region makes sense up to a cut-off region (dotted lines) and there is no left CFT.

chapter we are not interested in black holes formed by collapse, but rather in understanding the geometry dual to a *typical* black hole microstate in the CFT. This is defined as a pure state which is a random superposition of energy eigenstates

$$|\Psi_0\rangle = \sum_{E_i \in (E_0, E_0+\delta E)} c_i |E_i\rangle, \tag{3.1}$$

selected from a narrow energy band. Here c_i are random complex numbers, constrained to obey $\sum_i |c_i|^2 = 1$, and selected with the uniform Haar measure. We take

$$\delta E \sim R_{AdS}^{-1} O(N^0)\,, \qquad E_0 \sim R_{AdS}^{-1} O(N^2)\,. \tag{3.2}$$

Understanding the geometry of a typical black hole microstate is an important question, as — by definition— typical states represent the majority of black hole microstates of given energy. In addition, understanding the geometry of a typical state is important for the AdS/CFT version of the firewall paradox [12, 13, 16]. As mentioned above, the reader should keep in mind the difference between typical states and states formed by collapse. The number of states which can be formed by "reasonable" gravitational collapse is much smaller than those predicted by the Bekenstein-Hawking entropy, see for example an early discussion in [34]. Also, notice that strictly speaking the class of typical states defined above are not exactly the same as the late-time configuration of a collapsing black hole. For example, the standard inequality $\frac{1}{2}\left|\frac{d\langle A\rangle}{dt}\right| \leq \Delta A \Delta E$, where $\Delta A, \Delta E$ denote the variance of A and the energy respectively, implies that any state which undergoes gravitational collapse over a time-scale of the order of the AdS radius (hence there are observables A for which $\frac{d\langle A\rangle}{dt} \sim O(N^0)$) and which initially has a semi-classical description, i.e. $\Delta A \sim O(N^{-1})$,

must have an energy variance[2] which is $\Delta E \sim R_{AdS}^{-1} O(N)$. Such states are somewhat different from the typical states with narrow-energy band (3.2) that we consider here[3].

Typical states in the CFT look almost time-independent when probed by simple observables, which do not explicitly depend on time, since

$$\begin{aligned} \langle \Psi_0 | \frac{dA}{dt} | \Psi_0 \rangle = & \ \mathrm{Tr}[\rho_m \frac{dA}{dt}] + O(e^{-S}) \\ & = i \ \mathrm{Tr}[\rho_m [H, A]] + O(e^{-S}) = O(e^{-S}), \end{aligned} \tag{3.3}$$

where we introduced the microcanonical density matrix ρ_m relevant for the window $(E_0, E_0 + \delta E)$ and we used the approximation of a typical microstate to the microcanonical expectation value reviewed in subsection 3.4.1.1[4]. In the last equality we dropped the trace using $[\rho_m, H] = 0$. This suggests that the dual geometry to a typical pure state $|\Psi_0\rangle$ should be characterized by an approximate Killing isometry, which is timelike in the exterior region. A natural expectation within the AdS/CFT framework is that part of the dual geometry contains the exterior of a static black hole in AdS. For the benefit of the reader we summarise the relevant arguments in Appendix B.1. A natural question then is, does there exist an extension of this geometry behind the black hole horizon?

If the future horizon is smooth, then the dual geometry should contain at least part of the black hole interior. Since the ensemble of typical states is time-reversal invariant, we will conjecture that the dual geometry should also contain part of the white hole region. Finally, if the dual space-time contains parts of all these three regions, it is natural to assume that it should also contain part of the left asymptotic region. This leads us to the conjectured diagram in figure 3.3 for the dual geometry of a typical state. A typical state is in equilibrium so nothing is happening in it. Thus one may wonder what is the meaning of the statement that the dual geometry contains these regions. The operational meaning of this statement is that a class of perturbations of the boundary CFT can be described by low-energy effective field theory perturbations of the conjectured geometry. In other words, under a class of deformations the typical state responds "as if it had a smooth interior", partly extending into the left region. Notice that there is no left CFT, but rather the geometry is effectively inaccessible (and operationally meaningless) beyond the region indicated by the dotted lines, whose nature and location we will discuss later in section 3.1.3.

The left region of the conjectured geometry of a typical microstate is described by the "mirror operators", which we discuss below. The existence of a region, which is causally disconnected from the black hole exterior, is a consequence of the fact that the algebra $\mathcal{A}$ describing low-energy effective field theory experiments in the exterior has a nontrivial commutant $\mathcal{A}'$. Moreover, this commutant is entangled with $\mathcal{A}$. The fact that the geometry in the left region should be a "mirrored copy" of the right region, at least up to some cutoff on the far left, follows from the algebraic construction of the mirror operators (3.8). In that construction we will notice that the commutant $\mathcal{A}'$ is in some sense isomorphic to the original algebra. Combining together the small algebra $\mathcal{A}$ and its commutant $\mathcal{A}'$ we

[2] See appendix A of [25] for more details.

[3] It is interesting to better understand how collapsing black holes approximate certain classes of typical states at late times, and to clarify the role of complexity in studying the late time limit, see for example [35].

[4] Notice that to prove this we do not need to use the Eigenstate Thermalization Hypothesis (ETH) [36].

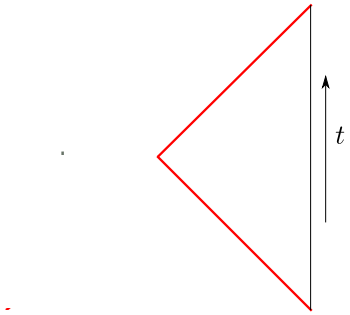


Figure 3.4: An alternative to our proposal: typical states have a firewall on the horizon. This would be consistent with the Killing isometry, but would correspond to modifications of general relativity at low curvatures.

get the black and white hole regions. We can think of the conjectured geometry dual to a typical state as a wormhole connecting the exterior of the black hole and the left interior region, which represents the space of the mirror operators. The operators $\mathcal{O}$ and $\widetilde{\mathcal{O}}$ are entangled in a similar way as the two sides of the thermofield state, and the emergence of the wormhole is reminiscent of the ER/EPR proposal [10]. The meaning of this proposed geometry is that we can use effective field theory on it to compute CFT correlators. These correlators can be localized within a finite time domain on the boundary, of the order of few scrambling times. In particular, the domain of validity of the conjectured diagram does not need to capture experiments extended over arbitrarily long time-scales. On the other hand, the finite time domain mentioned can be centered around any time, thus allowing us to access arbitrary regions in the proposed geometry.

It is important to consider what could be a possible alternative to the proposed geometry of figure 3.3. It is natural that the geometry will have to be consistent with the Killing isometry, at least in some approximate sense. One extreme possibility consistent with this symmetry is that the spacetime terminates on the past and future horizons, by a firewall or other object as indicated in figure 3.4. [5] This would violate our expectations from general relativity in a regime of low curvatures. If we want to avoid this scenario and if we want to preserve the smoothness of the horizon then we need to extend the geometry behind the horizon, up to some cutoff which must be consistent with the Killing isometry.

3.1.2 The Mirror Operators

The conjecture that a typical black hole state should be associated to a geometry, which has a smooth interior and moreover contains part of the left region follows from the construction of the "mirror operators", which was introduced in [6, 7, 8]. Similar conclusions from a somewhat different perspective were reached in [37]. The construction of the mirror operators starts by defining a small algebra of observables $\mathcal{A}$, which correspond to simple experiments described in effective field theory in the bulk. In a large N gauge theory $\mathcal{A}$ can be thought of as generated by products of a small number of single trace operators of low conformal dimension. Further details about the definition and limitation of the small algebra can be found in the references mentioned above. Here we only emphasize

[5] Here by "firewall" we broadly refer to any kind of structure at the location of the would-be-horizon, which would drastically modify the experience of the infalling observer from that predicted by general relativity.

that technically the set $\mathcal{A}$ is not a proper algebra, since we define it as the set of small products. However, in the large N limit this limitation is not important for what follows, so we will continue to refer to $\mathcal{A}$ as an algebra.

Given a typical black hole microstate $|\Psi_0\rangle$, we define the "small Hilbert space", also called code-subspace, as

$$\mathcal{H}_{|\Psi_0\rangle} = \text{span}\mathcal{A}|\Psi_0\rangle. \tag{3.4}$$

This subspace of the full CFT Hilbert space, contains all states that one can get starting from $|\Psi_0\rangle$ and acting on it with a small number of bulk operators. Hence, this subspace is the one relevant for describing effective field theory in the bulk[6]. An important algebraic property is that operators of the algebra $\mathcal{A}$ cannot annihilate the state $|\Psi_0\rangle$. This can be understood, for example, by noticing that for $A \in \mathcal{A}$ we have

$$||A|\Psi_0\rangle||^2 = \langle\Psi_0|A^\dagger A|\Psi_0\rangle = Z^{-1}\ \text{Tr}[e^{-\beta H}A^\dagger A] + O(1/S), \tag{3.5}$$

which is positive if we ignore the subleading corrections. Here we used the approximation of a typical pure state by the thermal ensemble, which is expected to be very good at least at leading order in large N.

The fact that $\mathcal{A}$ contains no annihilation operators for $|\Psi_0\rangle$ means that for the representation of the algebra $\mathcal{A}$ on the code subspace $\mathcal{H}_{|\Psi_0\rangle}$, the state $|\Psi_0\rangle$ is a *cyclic* and *separating* vector. As suggested by the Tomita-Takesaki theorem, see for example [38] for a review, this implies that the representation of the algebra $\mathcal{A}$ on the subspace $\mathcal{H}_{|\Psi_0\rangle}$ is reducible and the algebra has a non-trivial commutant $\mathcal{A}'$ acting on $\mathcal{H}_{|\Psi_0\rangle}$. The elements of $\mathcal{A}'$ are operators which commute with all elements of the algebra $\mathcal{A}$, and geometrically in the bulk correspond to local fields in a region which must be causally disconnected by the black hole exterior. It is natural to identify the region corresponding to $\mathcal{A}'$ with the left asymptotic region.

The commutant $\mathcal{A}'$ can be concretely identified by an analogue of the Tomita-Takesaki construction

$$SA|\Psi_0\rangle = A^\dagger|\Psi_0\rangle, \tag{3.6}$$

$$\Delta = S^\dagger S, \qquad S = J\Delta^{1/2}, \tag{3.7}$$

$$\widetilde{\mathcal{O}} = J\mathcal{O}J. \tag{3.8}$$

where J is an anti-unitary operator called *modular conjugation* and S is a general anti-linear map. Moreover, using large N factorization and the KMS condition relevant for equilibrium states, it is possible to show [8] that at large N the CFT Hamiltonian acts on the code subspace similar to the (full) modular Hamiltonian

$$\Delta = \exp[-\beta(H - E_0)] + O(1/N), \tag{3.9}$$

where we assume that the energy of $|\Psi_0\rangle$ is highly peaked around E_0.

To be more concrete, we will construct the small algebra $\mathcal{A}$ starting with single-trace operators in frequency space and later we discuss their Fourier transform to the time coordinate. The algebra $\mathcal{A}$ is defined as

$$\mathcal{A} = \text{span}\left\{\mathcal{O}_{\omega_1}, \mathcal{O}_{\omega_1}\mathcal{O}_{\omega_2}, ..., \mathcal{O}_{\omega_1}....\mathcal{O}_{\omega_n}\right\}, \tag{3.10}$$

[6] The code subspace knows about $1/N$ interactions which can be described within effective field theory.

where $n \ll N$. Of course this linear set is not a proper algebra as we demand $n \ll N$, for example we are not allowed to multiply together N single-trace operators $\mathcal{O}_{\omega_i}$. However, if n is very large (but much smaller than N) this set behaves approximately as an algebra for correlation functions involving a small number of operators. Moreover, the fact that $\mathcal{A}$ is not a proper algebra is important for the realization of the idea of black hole complementarity, as has been discussed in detail in [6, 8]. Nevertheless, we will continue referring to $\mathcal{A}$ as the "small algebra".

Now we clarify the nature of the Fourier modes generating the small algebra $\mathcal{A}$. We first consider the exact Fourier modes of operators, defined as

$$\mathcal{O}_{\omega}^{\text{exact}} \equiv \int_{-\infty}^{+\infty} dt\ e^{i\omega t}\mathcal{O}(t). \tag{3.11}$$

Usually ω takes values in $(-\infty, \infty)$. However, for the construction above we need to restrict the range of ω in some ways:

i) Since the spectrum of the dual CFT is assumed to be discrete[7], then for generic choice of real ω there will be no pair of states such that $E_i - E_j = \omega$. Hence, if ω is a generic real number, the operator $\mathcal{O}_{\omega}^{\text{exact}}$ defined by (3.11) will be zero. To avoid this we can bin together sets of frequencies in bins of size $\delta\omega$, which can be very small given that the typical energy level gap is $\beta^{-1}e^{-S}$. Therefore, we define the coarse-grained frequency operators

$$\mathcal{O}_{\omega} \equiv \frac{1}{\sqrt{\delta\omega}} \int_{\omega}^{\omega+\delta\omega} \mathcal{O}_{\omega'}^{\text{exact}} d\omega', \tag{3.12}$$

where now the set of allowed ω's is discretized with step $\delta\omega$. In (3.12) we have divided by $\sqrt{\delta\omega}$ in order to have an operator whose correlators are stable under small changes of the bin size $\delta\omega$. We will denote these coarse-grained Fourier modes simply as $\mathcal{O}_{\omega}$, without explicitly showing the choice of step $\delta\omega$, which is not important for most calculations. Alternatively to the binning procedure, we can think of the $\mathcal{O}_{\omega}$ in a distributional sense, where we always use these operators inside integrals over ω.

ii) We need to impose an upper cutoff in the allowed frequencies $|\omega| \leq \omega_*$. The reason is that the mirror operators are meaningful when the small algebra cannot annihilate the state. In a thermal state we find that $\langle \mathcal{O}_{\omega}^{\dagger}\mathcal{O}_{\omega}\rangle \propto e^{-\beta\omega}$. For large ω this is extremely close to zero, implying that the operator $\mathcal{O}_{\omega}$ almost annihilates the state. It is possible, and sufficient for our purposes, to take ω_* arbitrarily large but N-independent. We will discuss this and the possibility of scaling ω_* with N in subsection 3.1.3.

iii) If we want to describe the part of the black hole interior relevant for experiments initiated around some time t_0 in the CFT, then for the definition of the mirror operators we need to consider CFT operators which are localized only within a time band $t_0 \pm T_{\text{max}}$, where T_{max} is at least as large as several times the scrambling time. This means that the Fourier modes should be defined with respect to this IR-cutoff in time, which in turns effectively leads to a discretization of frequencies of order $\frac{1}{T_{\text{max}}}$. If T_{max} is large enough, this does not affect correlators significantly, see also the relevant discussion in [6]. Since our experiments are contained inside a time band given by T_{max} we do not probe the infinite past and infinite future of the state.

[7] For example consider the $\mathcal{N} = 4$ SYM on $\mathbb{S}^3 \times$ time.

From (3.8),(4.11) follows that at large N the mirror operators are defined by the equations[8]

$$\begin{aligned}\widetilde{\mathcal{O}}_\omega|\Psi_0\rangle &= e^{-\frac{\beta H}{2}}\mathcal{O}_\omega^\dagger e^{\frac{\beta H}{2}}|\Psi_0\rangle,\\ \widetilde{\mathcal{O}}_\omega\mathcal{O}_{\omega_1}...\mathcal{O}_{\omega_n}|\Psi_0\rangle &= \mathcal{O}_{\omega_1}...\mathcal{O}_{\omega_n}\widetilde{\mathcal{O}}_\omega|\Psi_0\rangle,\\ [H,\widetilde{\mathcal{O}}_\omega]\mathcal{O}_{\omega_1}...\mathcal{O}_{\omega_n}|\Psi_0\rangle &= \omega\widetilde{\mathcal{O}}_\omega\mathcal{O}_{\omega_1}...\mathcal{O}_{\omega_n}|\Psi_0\rangle.\end{aligned} \tag{3.13}$$

The last equation implies that the mirror operators are "gravitationally dressed" with respect to the CFT, i.e. the right boundary of the Penrose diagram. For the purposes of this chapter we extend these equations to define the mirror operators, even when we include $1/N$ effects. The extension of the definition of $\widetilde{\mathcal{O}}$ to subleading orders in $1/N$ is not unique, and for the reconstruction of local bulk fields more care about this issue should be taken. However, for the thought experiments we set up later, we will take (4.12) as the definition of the mirror operators even including $1/N$ corrections. One drawback of this choice is that it makes the hermiticity properties of the mirror operators somewhat more complicated, as explained in the paragraph below. On the other hand, this choice makes the comparison with the eternal black hole simpler, hence we will continue using it in this chapter.

We now come to the hermiticity properties of the mirror operators defined in (4.12). In particular from this definition it follows that to leading order at large N we have $\widetilde{\mathcal{O}}_{-\omega} = (\widetilde{\mathcal{O}}_\omega)^\dagger$, but this may no longer be true at subleading orders in $1/N$. To see that, consider the matrix elements of these operators on two general states in the small Hilbert space, which can be written as $A|\Psi_0\rangle$ and $B|\Psi_0\rangle$ with $A, B \in \mathcal{A}$. We have

$$\begin{aligned}\langle\Psi_0|A^\dagger\widetilde{\mathcal{O}}_{-\omega}B|\Psi_0\rangle &= e^{\frac{\beta\omega}{2}}\langle\Psi_0|A^\dagger B\mathcal{O}_\omega|\Psi_0\rangle,\\ \langle\Psi_0|A^\dagger(\widetilde{\mathcal{O}}_\omega)^\dagger B|\Psi_0\rangle &= e^{-\frac{\beta\omega}{2}}\langle\Psi_0|\mathcal{O}_\omega A^\dagger B|\Psi_0\rangle.\end{aligned} \tag{3.14}$$

In general it is not possible to argue that these two will be the same, which suggest that in general $\widetilde{\mathcal{O}}_{-\omega} \neq (\widetilde{\mathcal{O}}_\omega)^\dagger$. However, if we approximate correlators on typical pure states by thermal states, an approximation we will discuss extensively in section 3.4, we find

$$\begin{aligned}\langle\Psi_0|A^\dagger\widetilde{\mathcal{O}}_{-\omega}B|\Psi_0\rangle &= e^{\frac{\beta\omega}{2}}\mathrm{Tr}[\rho_\beta A^\dagger B\mathcal{O}_\omega] + O(1/N),\\ \langle\Psi_0|A^\dagger(\widetilde{\mathcal{O}}_\omega)^\dagger B|\Psi_0\rangle &= e^{-\frac{\beta\omega}{2}}\mathrm{Tr}[\rho_\beta\mathcal{O}_\omega A^\dagger B] + O(1/N).\end{aligned} \tag{3.15}$$

where ρ_β is the thermal density matrix. Finally we use the KMS condition for the thermal state, which implies $\mathrm{Tr}[\rho_\beta A^\dagger B\mathcal{O}_\omega] = e^{-\beta\omega}\mathrm{Tr}[\rho_\beta\mathcal{O}_\omega A^\dagger B]$, to conclude that within the code subspace we have

$$\widetilde{\mathcal{O}}_{-\omega} = (\widetilde{\mathcal{O}}_\omega)^\dagger + O(1/N) \tag{3.16}$$

In particular this means, for example, that the operator $\widetilde{\mathcal{O}}_\omega + \widetilde{\mathcal{O}}_{-\omega}$ is Hermitian up to $1/N$ corrections. We close the issue of hermiticity by saying that other extensions of the

[8] One might worry that conjugating $\mathcal{O}_\omega^\dagger$ with $e^{\pm\frac{\beta H}{2}}$ would result in a complicated operator, not necessarily obeying ETH. However, notice that the Fourier modes $\mathcal{O}_\omega^{\text{exact}}$ (3.11) obey precisely $[H,(\mathcal{O}_\omega^{\text{exact}})^\dagger] = \omega(\mathcal{O}_\omega^{\text{exact}})^\dagger$ and therefore $e^{-\frac{\beta H}{2}}(\mathcal{O}_\omega^{\text{exact}})^\dagger e^{\frac{\beta H}{2}} = e^{-\frac{\beta\omega}{2}}(\mathcal{O}_\omega^{\text{exact}})^\dagger$. After binning (3.12) we get an operator that obeys the ETH, assuming that $\mathcal{O}_\omega^\dagger$ did.

definition of mirror operators to subleading orders in $1/N$ have more manifest hermiticity properties and they may be more useful for other purposes[9].

Because of the restrictions in frequencies that we have imposed above, it is not meaningful to define the mirror operators for sharply localized operators $\mathcal{O}(t)$. As we will discuss later, in section 3.1.3, this is related to the fact that we do not expect to be able to reconstruct the entire left asymptotic region. We can still try to define approximately localized mirror operators $\widetilde{\mathcal{O}}(t)$ by using only the available Fourier modes, but the resulting operators will not behave as sharply localized operators. The approximation becomes better as we increase the cutoff ω_*.

We also emphasize that the equations (4.12) are supposed to hold only inside the code subspace. This means that the operators $\widetilde{\mathcal{O}}$ do not need to commute with the operators in $\mathcal{A}$ in an exact operator sense, but only when their commutator is inserted inside low-point correlation functions in the small Hilbert space. This is related to the idea of black hole complementarity.

3.1.2.1 Time Dependence of Mirror Operators

Since we will be considering time-dependent perturbations of the CFT Hamiltonian, it is necessary in this context to specify the CFT time when the mirror operators are applied, or at least specify a *time ordering* between them and also with the normal operators. Specifying the action of the operators on the small subspace $\mathcal{H}_{|\Psi_0\rangle}$, as in (4.12), is not sufficient to know the time when the operators act. This issue is discussed in more detail in Appendix B.2.

We will associate the mirror operators to physical time in the CFT in such a way that when we consider time-dependent perturbations of the CFT Hamiltonian with mirror operators, then the result can be described by effective field theory in the bulk. In terms of the conjectured Penrose diagram, the requirement of having a consistent effective field theory description in the bulk requires that as the CFT time t increases the corresponding mirror operators must move "upwards" on the left side of the diagram — the opposite choice would not lead to a globally consistent bulk causal structure. Imposing this condition we find a one-parameter family, labeled by T, of useful choices[10] for how to localize the mirrors in physical time. For each choice of T we define

$$\widetilde{\mathcal{O}}_T(t) = \int_{-\omega_*}^{\omega_*} d\omega \; e^{-i\omega(t-T)} \widetilde{\mathcal{O}}_\omega, \tag{3.17}$$

where t labels the physical CFT time at which the operator is localized. These time-dependent real-space operators are the Fourier transform of a function which has i) an upper cutoff in ω and ii) the frequencies are discretized. The discretization of frequencies does not impose a serious restriction at time-scales of $O(1)$, given that $\delta\omega$ can be very small. On the other hand the upper cutoff ω_* means that we should not be trying to resolve time with resolution smaller than $\Delta t \sim \frac{1}{\omega_*}$.

[9] For example, if we define the mirror operators by using the Tomita-Takesaki equations (3.6)-(3.8) to all orders in $1/N$, and without making the approximation (4.11), then it can be shown that: if $\mathcal{O}$ is Hermitian, then $\widetilde{\mathcal{O}}$ is Hermitian. On the other hand, other approximations at subleading order in $1/N$ become harder with this definition.

[10] The sense in which this way of localizing the mirrors in time is useful, is that active perturbation using these mirrors can be represented geometrically by a space-time diagram and effective field theory on it. See also appendix B.2.

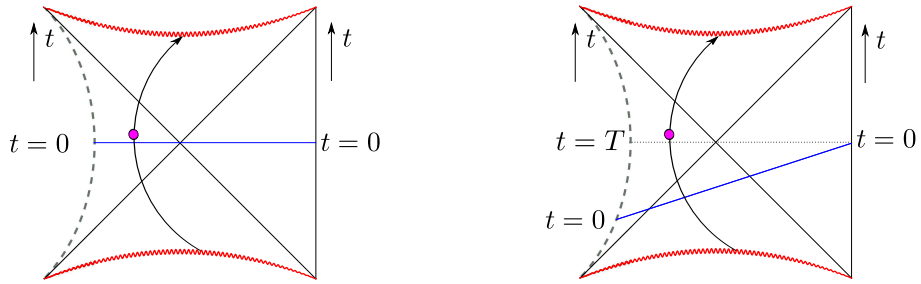


Figure 3.5: Two different choices of localizing the mirror operators in physical time, corresponding to different identifications of physical time t with the coordinate on the left region. The two diagrams correspond to different choices of T, the left diagram is for $T = 0$. Notice that the bulk geometry of the left region, if it is defined relationally to the right boundary, is the same in both cases. However, the types of allowed active perturbations of the state are different.

Time-dependent perturbations of the Hamiltonian by mirror operators have a simple bulk interpretation if the same choice of T is used for all mirror operators in the calculation. It does not matter what T is, but one should not mix mirror operators with different choices of T, or else the bulk dual would be complicated. Notice that the choice of T can also be understood in terms of representing the typical microstate in terms of the "time-shifted eternal black hole" [15], where we move one of the two boundaries by a large diffeomorphism corresponding to time translation by T.

We think of the operators $\widetilde{\mathcal{O}}_T(t)$ as acting at physical CFT time t. This means that, for given and fixed choice of T, different mirror operators are time-ordered according to the obvious way with respect to the parameter t. i.e. a perturbation by $\widetilde{\mathcal{O}}_T(t_1)$ can affect operators $\widetilde{\mathcal{O}}_T(t_2)$ if $t_2 \geq t_1$. Notice the important minus sign in the exponential in (4.13), relative to what one would naively expect for the inverse Fourier transform in the conventions of (3.11) and from the fact that $\widetilde{\mathcal{O}}$ seems to be associated to $\mathcal{O}_\omega^\dagger$ from the first equation of (4.12) . This minus sign can be thought of effectively as a time reversal around $t = T$. This minus sign also implies that in the Heisenberg picture, these operators obey

$$\widetilde{\mathcal{O}}_T(t) = e^{-iHt}\widetilde{\mathcal{O}}_T(0)e^{iHt}. \tag{3.18}$$

The fact that this is the opposite evolution than the usual one implies that if we think of them as operators in the Schroedinger picture they are explicitly time-dependent, in such a way that effectively they behave as if they were running backwards in time i.e. in Schroedinger picture with physical time t we would have the *explicit* time-dependence

$$[\widetilde{\mathcal{O}}_T(t)]_{\rm S} = e^{-2iHt}[\widetilde{\mathcal{O}}_T(0)]_{\rm S}e^{2iHt}. \tag{3.19}$$

In the rest of this paper we will be working with the choice $T = 0$. For notational ease we will write $\widetilde{\mathcal{O}}_{T=0} = \widetilde{\mathcal{O}}$.

Finally, as we discussed in the previous section, and in particular equation (3.16), the operator $\widetilde{\mathcal{O}}_T(t)$ is not Hermitian at subleading orders in $1/N$. Hence if we want to use it as a perturbation to the CFT Hamiltonian we need to add subleading $1/N$ corrections to the operator, to ensure that the perturbed Hamiltonian is Hermitian. Provided that typical state correlators can be well approximated by thermal correlators at large N, a condition

that we will formulate as a technical conjecture in section 3.4, these subleading corrections necessary to promote (4.13) into a Hermitian operator will not alter the relevant results to leading order. For the rest of this chapter whenever we discuss perturbations involving $\widetilde{\mathcal{O}}$ we will always assume that we have dressed up the operator so that it is Hermitian.

3.1.3 On the Boundary of the Left Region

Since we are considering the geometry dual to a single CFT, we do not expect to be able to reconstruct the left region of the Penrose diagram all the way towards asymptotic AdS infinity. This is related to the fact that we are not able to define the mirror operators for sharply localized operators in time, or equivalently it is related to the cutoff frequency ω_* that we introduced in section 3.1.2. If we think of a wavepacket in the background of an AdS Schwarzchild black hole, then if we want to localize the packet close to the boundary we need to use high frequencies ω. If we have a cutoff in the allowed frequencies $|\omega| < \omega_*$ then any wavepacket constructed with this cutoff will have limited reach towards the boundary. In the limit of large ω_* we find that this translates into a cutoff in the usual r-coordinate in global AdS[11] as

$$r < r_* \qquad , \qquad r_* \sim \omega_*. \tag{3.20}$$

Here we are working in units where $R_{\rm AdS} = 1$. This estimate follows from the gravitational potential of AdS and from analyzing what frequencies are necessary in order to localize wavepackets around a particular region of r. Hence, the question of how far towards the left we can extend the geometry depends on the cutoff ω_*. First we start with a conservative estimate: if we take the large N limit, we can take ω_* to be as large as we like, provided that it is not N-dependent. This also means that the cutoff r_* can be arbitrarily large, though N-independent. In particular this means that the left geometry can be extended to (arbitrarily) many times the Schwarzschild radius of the black hole towards the left. This is sufficient to formulate most of the thought experiments that we want to consider.

Of course it is interesting to understand how far the left region extends. In order for the geometry to be operationally meaningful, a probe must be able to explore it. Any probe in the left region should be thought of as a spontaneous out of equilibrium excitation "borrowing" energy from the black hole. Hence the black hole mass provides an upper limit to the energy that these probes can have. Taking into account the redshift factor near the AdS boundary we find that this implies an ultimate upper bound $r_* \sim M$ or $r_* \sim O(N^2)$ — but the actual bound may be much smaller. This is consistent with the fact that the mirror operators can not be defined for frequencies of order $\omega \sim O(N^2)$, since $\langle \mathcal{O}^\dagger_\omega \mathcal{O}_\omega \rangle \sim e^{-\beta\omega}$ is almost zero, see discussion in section 3.1.2. So far we have identified that the left region can be reconstructed *at least* up to $r_* \sim \alpha$ where α can be arbitrarily large but N independent, and *at most* up to $r_* \sim O(N^2)$. The actual cutoff region must lie somewhere in-between. We have not been able to identify the more precise limit of the bulk reconstruction but this is clearly a very interesting question. We would also like to pose the following question: what happens to the space-time when we approach this cutoff region? Our conjecture is that there is no breakdown of effective field theory anywhere in the left region, but the limitation of the reconstruction arises simply from the fact that

[11] These are coordinates where empty global AdS would have the form $ds^2 = -(1+r^2)dt^2 + \frac{1}{1+r^2}dr^2 + r^2 d\Omega^2_{d-1}$.

there is a restriction in the energies of the allowed probes moving in the left region. In particular, the energies of the allowed probes are bounded and this bound is what makes it impossible, even in principle, to probe the far-left region of the Penrose diagram and not a breakdown of bulk effective field theory. This situation is very different from that of non-typical states. For a class of such states, it was suggested that the bulk geometry has a left region bounded by some kind of end-of-the-world membrane [24, 31], which plays the role of a hard cutoff.

3.1.4 Comments on the Hamiltonian

Let us call M the ADM mass as measured in the bulk from the right side of the black hole. We argued above that the effective cutoff on the left can be pushed quite far when we are working in the large N limit. Hence it is natural to define an analogue of the left ADM mass $\widetilde{M}$. The first law [39] applied to the two-sided Cauchy slice Σ up to the left cutoff implies

$$\delta M - \delta\widetilde{M} = \delta K^{\text{full}}_{\text{bulk}}, \tag{3.21}$$

where $K^{\text{full}}_{\text{bulk}} = \int_\Sigma *(\xi T_{\text{bulk}})$, ξ is the Killing vector field and T_{bulk} is the bulk stress tensor corresponding to EFT excitations in the left and right regions. The quantity $K^{\text{full}}_{\text{bulk}}$ can naturally be split into the right and left contributions $K^{\text{full}}_{\text{bulk}} = K - \widetilde{K}$. Since we consider the $\widetilde{\mathcal{O}}$ operators to be gravitationally dressed with respect to the right, we have $\delta\widetilde{M} = 0$. This means that in the small Hilbert space the CFT Hamiltonian acts as

$$H = M = E_0 + K^{\text{full}}_{\text{bulk}}, \tag{3.22}$$

where E_0 is the energy of $|\Psi_0\rangle$.

We notice that according to the identification (3.22) excitations which are created in the left region by right-dressed operators have negative energy with respect to the CFT Hamiltonian. We provide a perhaps pedagogically more direct demonstration of the negative energy of excitations in the left region by considering a particular class of perturbations in appendix B.3. This negative energy is also related to the following point: the "physical time", i.e. the time ordering, for the left region is taken to be pointing upwards in the Penrose diagram. On the other hand taking commutators with the CFT Hamiltonian moves the points downwards. In other words the geometric action of H coincides with the Killing vector field. This means that "physical time evolution" in the mirror region is not generated by H but rather by $-H$. The reason this happens is that if we think of the mirror operators as being localized in time according to the rule of the previous subsections, then these mirrors are actually explicitly time-dependent operators, as discussed around equation (3.19). Hence their physical time evolution in the Heisenberg picture is not given by $\frac{d\widetilde{\mathcal{O}}}{dt} \stackrel{?}{=} i[H, \widetilde{\mathcal{O}}]$, but rather $\frac{d\widetilde{\mathcal{O}}}{dt} = i[H, \widetilde{\mathcal{O}}] + \frac{\partial\widetilde{\mathcal{O}}}{\partial t} = -i[H, \widetilde{\mathcal{O}}]$, since the explicit time dependence in the Schröedinger picture has to be taken into account.

3.1.5 Perturbations of Typical States

In this chapter, we only analyze small perturbations of the quantum fields on top of the background geometry. These correspond to excitations which change the CFT energy by factors of $O(N^0)$.

Typical states are closely related to equilibrium states, defined as states on which simple correlators are almost time-independent. In the rest of this chapter we will be discussing

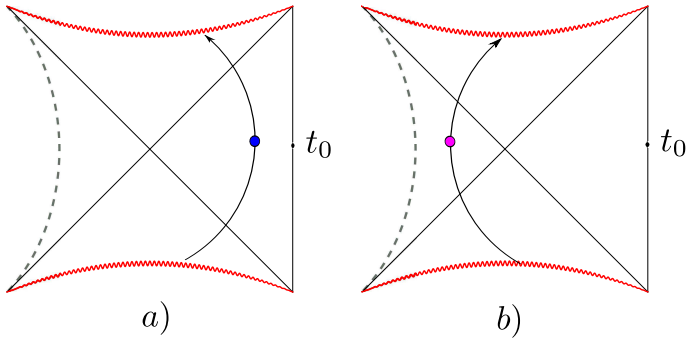


Figure 3.6: Autonomous non-equilibrium states of the form: a) $U(\mathcal{O}(t_0))|\Psi_0\rangle$ and b)$U(\widetilde{\mathcal{O}}(t_0))|\Psi_0\rangle$.

various perturbations of a typical state $|\Psi_0\rangle$. These perturbations can either be thought of as excited autonomous states, or as states where we actively perturb the system by turning on sources, or combinations of the two. Here we present some examples.

3.1.5.1 Autonomous Excited States

We use the term "autonomous states" to refer to quantum states where the Hamiltonian of the theory is not modified as a function of time. Hence, the entire history of the state is given by time evolution with respect to $H_{\rm CFT}$ and we are computing correlators on that state. This has to be contrasted with "actively-perturbed states", where we modify the CFT Hamiltonian for some period of time.

In AdS/CFT for autonomous states we do not turn on any sources on the boundary. Thus states are given as initial conditions and evolve with a time independent hamiltonian after that, for example

$$|\Psi\rangle = U(\mathcal{O}(t_0))|\Psi_0\rangle. \tag{3.23}$$

where $|\Psi_0\rangle$ is a typical state. The state $|\Psi\rangle$ can be thought of as a state which was prepared to undergo a spontaneous fluctuation out of equilibrium at around $t = t_0$. The unitary could be something of the form $e^{ig\mathcal{O}(t_0)}$, appropriately smeared. For $t \ll t_0$ the state looks like an equilibrium state. At around $t = t_0$ an excitation seems to be emitted from the past horizon, coming from the white hole region, reaches a maximum distance in AdS and falls back into the future horizon. The difference of energy between $|\Psi\rangle$ and $|\Psi_0\rangle$ is

$$\begin{aligned}\Delta E = \langle\Psi|H|\Psi\rangle - \langle\Psi_0|H|\Psi_0\rangle = \langle\Psi_0|U^\dagger[H,U]|\Psi_0\rangle = {\rm Tr}[\rho_\beta U^\dagger[H,U]] + O(1/S)\\ = {\rm Tr}[U\rho_\beta U^\dagger H] - {\rm Tr}[\rho_\beta H] + O(1/S),\end{aligned} \tag{3.24}$$

where $\rho_\beta = \frac{e^{-\beta H}}{Z}$ and we used the approximation of a typical state by a thermal ensemble. This way of organizing the computation aims at keeping the error terms at $O(1/S)$. If we ignore the $1/S$ corrections, and use the positivity of the relative entropy, we find that $\Delta E \geq 0$. To see that, we consider $S_{\rm rel}(\rho|\sigma)$ for $\sigma = \rho_\beta$ and $\rho = U\rho_\beta U^\dagger$. We have $S_{\rm rel} = \Delta K - \Delta S \geq 0$. For these two density matrices we have $\Delta S = 0$ hence $\Delta K = {\rm Tr}[U\rho_\beta U^\dagger H] - {\rm Tr}[\rho_\beta H] \geq 0$, or $\Delta E \geq 0$.

Another example is

$$|\Psi\rangle = U(\widetilde{\mathcal{O}}(t_0))|\Psi_0\rangle. \tag{3.25}$$

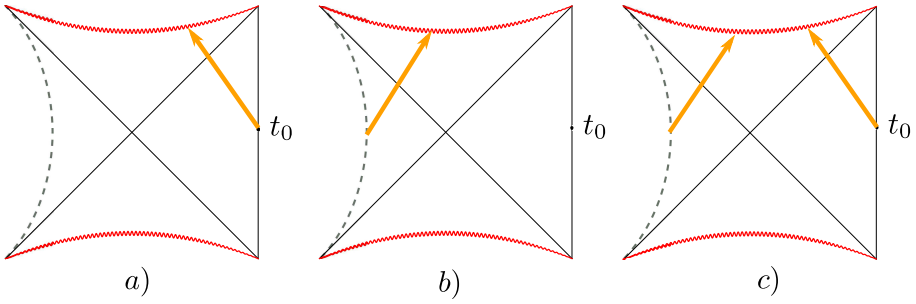


Figure 3.7: a) A usual quench b) A mirror quench c) A mixed quench: the analogue of a double-trace perturbation

These are states which are prepared to undergo a spontaneous fluctuation out of equilibrium at around $t = t_0$, but now in the space of mirrors. The two types of autonomous non-equilibrium states that we have already discussed are schematically depicted in figure 3.6.

The states (3.25) can also be written as

$$|\Psi\rangle = W|\Psi_0\rangle. \tag{3.26}$$

where

$$W \equiv e^{-\frac{\beta H}{2}} U(\mathcal{O}(-t_0)) e^{\frac{\beta H}{2}}. \tag{3.27}$$

Notice that while W is not a unitary, we have $\langle\Psi_0|W^\dagger W|\Psi_0\rangle = 1 + O(1/S)$, see [25] for more details. If we now estimate the leading order change of the energy we find

$$\begin{aligned} \Delta E &= \langle\Psi|H|\Psi\rangle - \langle\Psi_0|H|\Psi_0\rangle = \langle\Psi_0|W^\dagger[H,W]|\Psi_0\rangle + O(1/S) \\ &= \mathrm{Tr}[\rho_\beta W^\dagger[H,W]] + O(1/S) = -\mathrm{Tr}[U^\dagger\rho_\beta U H] + \mathrm{Tr}[\rho_\beta H] + O(1/S). \end{aligned} \tag{3.28}$$

It is interesting that we now find $\Delta E \leq 0$. This is consistent with the discussion of the previous subsection, where we argued that placing excitations in the left region lowers the energy of the CFT. This may seem a little surprising, as we are arguing that the fixed operator $e^{-\frac{\beta H}{2}} U(\mathcal{O}(t_0)) e^{\frac{\beta H}{2}}$ can lower the energy of a *typical* state, which at first sight seems to be inconsistent with the fact that there are fewer states at lower energies. The resolution of this apparent puzzle was discussed in detail in [25]. We provide a more explicit example of how this works in the SYK model in subsection 3.3.1.1.

Of course we can also consider more general perturbations involving combinations of excitations in both regions.

3.1.5.2 Perturbations of the Hamiltonian

We can also consider states where the Hamiltonian is perturbed at some particular moment in time, for example

$$H(t) = H_0 + f(t)A(t), \tag{3.29}$$

where $A(t)$ is Hermitian operator localized at time t and $f(t)$ is some smearing function, peaked around some particular t_0. The operator A can be made out of the $\mathcal{O}$'s or the $\widetilde{\mathcal{O}}$'s or both, but it is important that all the constituents of A are localized at the physical time

t. The most familiar type of perturbation is to take $A(t)$ to be a simple operator made out of local CFT operators $\mathcal{O}(t)$. This injects some energy into AdS from the boundary. In some approximation it can be described by a shockwave falling into an AdS black hole, as shown in figure 3.7.

If A is a sharply localized local operator, then in some sense the excitations are created near the boundary of AdS. We could also consider smeared A's, for example using the HKLL construction, so that the excitations can be created at some finite depth in AdS. The states that are produced by the time-dependent perturbation (3.29) have the property that at time $t > t_0$ they look like the corresponding autonomous states (provided A has been smeared), while for $t < t_0$ they look like equilibrium states.

As we discussed before, the boundary perturbation increases the energy of the state. In a shockwave approximation the spacetime before the perturbation has mass M while after the perturbation it has mass $M+\delta M$. The matching conditions across the shockwave relate δM to the stress tensor on the shockwave. We can consider excitations created in the left region,

$$H(t) = H_0 + f(t)\widetilde{A}(t), \tag{3.30}$$

by perturbing the Hamiltonian with an operator $\widetilde{A}$ constructed out of the mirror operators $\widetilde{\mathcal{O}}$[12]. In the shockwave approximation we see the spacetime diagram in 3.7. For a consistent bulk effective field theory interpretation it is important to use the specific precursors of the mirrors which localize them at the appropriate value of physical time. Notice that since the operators $\widetilde{\mathcal{O}}$ are gravitationally dressed with respect to the right, an active perturbation of the CFT Hamiltonian by $\widetilde{\mathcal{O}}$ will introduce a gravitational Wilson line in the bulk extending from the right boundary all the way into the left region where the shockwave seems to originate, see for example [40]. At order $O(1/N^2)$ these gravitational Wilson lines will backreact on the geometry and their effect has to be included as contributing to the Einstein equations. Understanding the effect of these Wilson lines on the trajectories of probes in the bulk is an interesting question, which we discuss in some more detail in subsection 3.2.2.3, but we postpone a more complete analysis to future work.

It is interesting to consider the backreaction of the shockwave. Since the operators are gravitationally dressed with respect to the right, the left-mass of the spacetime does not change. On the other hand the right mass below the Wilson lines will be M while above the Wilson lines $M+\delta M$. As we saw before, the effect of perturbing the CFT Hamiltonian by $\widetilde{\mathcal{O}}$ leads to $\delta M < 0$, which corresponds to *lowering* the CFT energy. We remind that this is not inconsistent with the 2nd law of thermodynamics, because the perturbation is state-dependent.

Finally we can consider perturbations $A(t)$ which are made out of both $\mathcal{O}$'s and their mirrors. It is important that these operators are localized at the same physical time. We will be interested in the particular class of perturbations of the schematic form

$$H(t) = H_0 + f(t)\mathcal{O}(t)\widetilde{\mathcal{O}}(t). \tag{3.31}$$

This produces two shockwaves as indicated in the figure 3.7 and as we will argue is the 1-sided analogue of the double-trace perturbations introduced in [5]. We will discuss this type of perturbation in more detail in section 3.2.

[12] As discussed before, we should make sure that this is a Hermitian operator by adding appropriate $1/N$ corrections.

3.2 Traversable one-sided black holes

In this section we will discuss in detail the state-dependent perturbation of the class (3.31), which we write more precisely as

$$H(t) \equiv H_0 + gV(t) = H_0 + gf(t)\,\mathcal{O}(t)\,\widetilde{\mathcal{O}}(t), \tag{3.32}$$

We assume that originally the CFT is in a typical pure state $|\Psi_0\rangle$. Here $\widetilde{\mathcal{O}}$ are the mirror operators defined by (4.12) with respect to $|\Psi_0\rangle$. As discussed before, we implicitly assume that these operators are supplemented by appropriate $1/N$ corrections to make them Hermitian. In the expression above we think of the operators as in the interaction picture. The function $f(t)$ is taken to be highly peaked around some time, say $t = 0$. Here and hereafter, we assume that the simple operators $\mathcal{O}(t)$ as well as the mirror operators $\widetilde{\mathcal{O}}(t)$ have uniform support over the entire space domain

$$\mathcal{O}(t) \equiv \int d^{d-1}\,\mathbf{x}\,\mathcal{O}(t,\mathbf{x}). \tag{3.33}$$

We could also consider generalizations where the perturbation uses many light operators $\sum_{i=1}^{K} f(t)\,\mathcal{O}_i(t)\,\widetilde{\mathcal{O}}_i(t)$, which simplifies some computations at large K [9].

We discuss some details of the operator V in subsection 3.2.2. Our goal in this section is to use properties of the typical state $|\Psi_0\rangle$ and the operator V to provide evidence that typical states in the CFT correspond to the geometry proposed in figure 3.1. In particular, this will be evidence that typical black holes in AdS have a smooth interior and a left-exterior region with an effective cutoff, as depicted in Penrose diagram 3.1. Our analysis will involve doing two thought experiments in the CFT.

- In **Experiment 1**, we imagine sending a probe made from a mirror particle $\widetilde{\phi}$ at time $-t_*$, which is of the order of scrambling time $\frac{\beta}{2\pi}\log(S)$. Then we turn on the perturbation V to the Hamiltonian at time $t = 0$. Finally we compute the expectation value of a simple operator in the CFT on the resulting state. If there is a signal of the expected form, that would imply that the mirror particle escaped the horizon. This experiment is depicted in figure 3.8 and described in subsection 3.2.3.

- In **Experiment 2**, we throw in a particle ϕ into the black hole at time $t = -t_*$. The perturbation V is then turned on at time $t = 0$. We then compute the response of the signal ϕ on a mirror operator $\widetilde{\phi}$ at time $t = t_*$. A non-vanishing response implies the ability of a boundary observer to reconstruct a message thrown into the black hole using the mirror particles. This experiment is reminiscent of the Hayden-Preskill protocol [30]. We will elaborate on this connection in section 3.5. This experiment is depicted in figure 3.9 and described in subsection 3.2.3.2.

Both experiments involve calculating out-of-time-ordered correlators in the typical state $|\Psi_0\rangle$. In fact, using the defining properties of the mirror operators (4.12), we will show that these correlators are approximately equal to left-right correlators with the same structure in the two-sided black hole geometry perturbed by a double-trace operator. We will discuss the errors appearing in this approximation, which are small in the large N limit. The left-right correlators in the two-sided black hole geometry were analyzed in [5] and [9]. We will thus review their calculation first.

3.2.1 Double-Trace Perturbation of the Two-Sided Black Hole

We will now review some aspects of the works [5, 9], where it was argued that a time-dependent double-trace perturbation to the two-sided black hole makes the wormhole connecting the two sides traversable. The eternal two-sided black hole is dual to the thermofield double state (TFD) which is

$$|\Psi_{\text{TFD}}\rangle = \frac{1}{Z(\beta)^{1/2}} \sum_E e^{-\frac{\beta E}{2}} |E\rangle_L |E\rangle_R , \tag{3.34}$$

where the sum is over all energy eigenstates, and β is the inverse temperature of the black hole.

No information can be transferred between the left and the right CFT, since the operators on the left and right commute $[\mathcal{O}_R, \mathcal{O}_L] = 0$. However, coupling the two CFTs with a double-trace perturbation as

$$H(t) = H_0 + g\, f(t)\, \mathcal{O}_R(t)\, \mathcal{O}_L(t), \tag{3.35}$$

allows for transfer of information between the two CFTs. Here, $\mathcal{O}_R, \mathcal{O}_L$ are again integrated over all of space. In [9] the perturbation was written as a sum over many fields. In principle we would compute the effect of this perturbation in the interaction picture by a time-ordered exponential

$$\mathcal{U}_g = \mathcal{T} e^{ig \int dt' f(t') \mathcal{O}_L(t')\, \mathcal{O}_R(t')}, \tag{3.36}$$

The transfer of information between the CFTs can be diagnosed using the correlator

$$\mathcal{C} \equiv \langle \Psi_{\text{TFD}}|\, e^{-i\epsilon\phi_L(-t_*)}\, \mathcal{U}_g^\dagger\, \phi_R(t) \mathcal{U}_g\, e^{i\epsilon\phi_L(-t_*)}\, |\Psi_{\text{TFD}}\rangle , \tag{3.37}$$

which is the one-point function of a field $\phi_R(t)$ in the right exterior region sourced by a field $\phi_L(-t_*)$ in the left exterior region, in the presence of the double-trace perturbation. Here t_* is of the order of the scrambling time $\frac{\beta}{2\pi} \log S$. For appropriate sign of g it was shown in [5, 9] that

$$\mathcal{C} \neq 0, \tag{3.38}$$

indicating information transfer.

In the bulk, the double-trace perturbation can be thought of as inserting $\mathcal{O}(N^0)$ energy into the bulk which propagates almost lightlike and thus represents two shockwaves falling into the black hole, from both of the boundaries. One can also find a post-perturbation geometry that is smoothly glued along these shockwaves. These shockwaves backreact on the eternal black hole geometry such that the IR of the geometry is changed [5]. In particular, for an appropriate sign of g, the quantum stress energy tensor of these shockwaves violates the averaged null energy condition, and thus allows for the wormhole to become traversable. The traversability can be seen by a non-zero commutator of two matter fields in the left and the right exterior region, captured for example by the correlator (3.37), thus allowing for transfer of information between the two boundaries.

It is instructive to discuss the quantum stress energy tensor in some more detail. For a scalar bulk field with mass m in the right exterior region, it is given by

$$T_{\mu\nu}(x) = \partial_\mu \phi(x)\, \partial_\nu \phi(x) - \frac{1}{2}\, g_{\mu\nu}\, \partial^\rho \phi(x)\, \partial_\rho \phi(x) - \frac{1}{2}\, g_{\mu\nu}\, m^2\, \phi^2(x). \tag{3.39}$$

Its expectation value in the state perturbed by (3.35) can be calculated using the point-splitting method

$$\langle T_{\mu\nu}\rangle = \lim_{x'\to x}\left[\partial_\mu\,\partial'_\nu\,G(x,x') - \frac{1}{2}\,g_{\mu\nu}\,g^{\rho\sigma}\,\partial_\rho\,\partial'_\sigma\,G(x,x') - \frac{1}{2}\,g_{\mu\nu}\,m^2\,G(x,x')\right], \tag{3.40}$$

where the short distance singularities have to be subtracted. One then only needs to know the corrected bulk two-point function $G(x,x')$ of field $\phi(x)$ with itself, which can be computed in perturbation theory in g [5]

$$\begin{aligned} G(x,x') &\equiv \langle\Psi_{\rm TFD}|\,\mathcal{U}_g^\dagger\,\phi(x)\,\phi(x')\,\mathcal{U}_g\,|\Psi_{\rm TFD}\rangle \\ &= G_0(x,x') - ig\bigg(\int_{t_0}^{t} dt_1\,f(t_1)\,\langle\Psi_{\rm TFD}|\left[\mathcal{O}_R(t_1)\,\mathcal{O}_L(t_1),\phi(t)\right]\phi(t')\,|\Psi_{\rm TFD}\rangle \\ &\quad + \int_{t_0}^{t'} dt_2\,f(t_2)\,\langle\Psi_{\rm TFD}|\,\phi(t)\left[\mathcal{O}_R(t_2)\,\mathcal{O}_L(t_2),\phi(t')\right]\,|\Psi_{\rm TFD}\rangle\bigg) + O(g^2), \end{aligned} \tag{3.41}$$

where t_0 is the time before which the function $f(t)$ vanishes, G_0 denotes the two-point function in the absence of the perturbation and we have suppressed space coordinates. This can be further simplified using $[\phi,\mathcal{O}_L]=0$ since we assumed that ϕ is in the right region. For the calculation of the $O(g)$ term the entanglement between the two CFTs plays a crucial role. If the two CFTs were in a state very different from the TFD state, firstly there would be no wormhole in the bulk and secondly a simple double-trace perturbation would not lead to a drastic modification of the bulk two-point function.

For the wormhole to be traversable, certain no-go theorems of semi-classical gravity need to be avoided. These often use the average of the local energy, which is

$$\int_{-\infty}^{\infty} du\,\langle T_{uu}\rangle \equiv \int_{u=-\infty}^{u=\infty} du\,\langle T_{\mu\nu}\,K^\mu\,K^\nu\rangle, \tag{3.42}$$

where the null coordinate u runs along the semi-infinite null geodesic very close to the horizon and K^μ denotes a unit vector tangent to it. It was checked in [5] that the zeroth order term in g in (3.40) coming from $G_0(x,x')$ integrates to zero, as expected. At the first subleading order, i.e. at $\mathcal{O}(g)$, we already see that the averaged null energy is proportional to g, with the proportionality function being the null integral over derivatives of the subleading two-point function. It was shown in [5] that for appropriate choice of g we have

$$\int_{-\infty}^{\infty} du\,\langle T_{uu}\,\rangle < 0. \tag{3.43}$$

This shows that the wormhole can be made traversable.

3.2.2 State-Dependent Perturbations in a Single CFT

From now on we consider a single CFT. The typical, heavy pure state in the CFT

$$|\Psi_0\rangle = \sum_{E_i\in(E_0,E_0+\delta E)} c_i|E_i\rangle, \tag{3.44}$$

is a microstate of a typical large black hole in AdS with one asymptotic boundary. Consider the time-dependent perturbation to the CFT Hamiltonian

$$H(t) \equiv H_0 + g\,V(t) = H_0 + g\;f(t)\,\mathcal{O}(t)\,\widetilde{\mathcal{O}}(t), \tag{3.45}$$

where $\mathcal{O}$ is a simple operator and $\widetilde{\mathcal{O}}$ is its mirror defined in subsection 3.1.2. As discussed in the beginning of section 3.2 the smearing function $f(t)$ is assumed to be highly peaked around $t = 0$. The operators are uniformly smeared on the spatial sphere on which the CFT is defined. Remember that the operators $\widetilde{\mathcal{O}}(t)$ have been defined so that they contain frequencies only up to some cut-off ω_*. We similarly define the operators $\mathcal{O}(t)$ to be somewhat smeared in time, so that they also contain frequencies up to ω_*. As in [9], we can also consider perturbations involving a sum over many different pairs of operators of the form $\sum_i \mathcal{O}_i\widetilde{\mathcal{O}}_i$.

3.2.2.1 Energy Change After the Perturbation

A natural diagnostic to study after the perturbation is the change in the energy of the typical state $|\Psi_0\rangle$. The total energy of the state after the perturbation is given by

$$E \equiv \langle\Psi_0|\,\widehat{\mathcal{U}}_g^\dagger\, H_0\,\widehat{\mathcal{U}}_g\,|\Psi_0\rangle\,, \tag{3.46}$$

where

$$\widehat{\mathcal{U}}_g = \mathcal{T}e^{ig\int dt' f(t')\mathcal{O}(t')\,\widetilde{\mathcal{O}}(t')}. \tag{3.47}$$

The energy before the perturbation

$$E_0 \equiv \langle\Psi_0|\,H_0\,|\Psi_0\rangle\,, \tag{3.48}$$

is fixed in terms of the coefficients c_i and the eigenvalues E_i that define the typical state $|\Psi_0\rangle$. Expanding up to first order in g we find

$$\Delta E = \int dt' f(t')\,\langle\Psi_0|\; ig\big[H_0, \mathcal{O}(t')\widetilde{\mathcal{O}}(t')\big]|\Psi_0\rangle + O(g^2). \tag{3.49}$$

It is easy to see that the first order term is zero. We consider the two-point function

$$G(t_1,t_2) = \langle\Psi_0|\,\mathcal{O}(t_1)\widetilde{\mathcal{O}}(t_2)\,|\Psi_0\rangle\,. \tag{3.50}$$

Using equations (4.12) and (4.13), this two-point function can be shown to be a function only of $t_1 + t_2$. Then the $O(g)$ term in equation (3.49) simplifies to

$$\Delta E_1 \sim (\partial_{t_1} - \partial_{t_2})G(t_1+t_2)|_{t_1=t_2} = 0, \tag{3.51}$$

where we used the explicit time-dependence of the mirror operators, which gives the minus sign in front of ∂_{t_2}.

It is instructive to compare this to the change in energy of the thermofield double state after a double-trace perturbation $e^{ig\,\mathcal{O}_L\,\mathcal{O}_R}$. Even the order $\mathcal{O}(g)$ change in energy is non-zero [5], as one sees from the non-zero value for the correlator

$$\begin{aligned}\Delta E_1^{\text{TFD}} &= -ig\;\langle\Psi_{\text{TFD}}|\;[\mathcal{O}_R, H_R]\,\mathcal{O}_L\;|\Psi_{\text{TFD}}\rangle \\ &= -\frac{ig}{Z(\beta)}\,\text{Tr}\Big(e^{-\beta H_R}\,[\mathcal{O}_R, H_R]\,e^{-\frac{\beta H_R}{2}}\,\mathcal{O}_R^\dagger e^{\frac{\beta H_R}{2}}\Big),\end{aligned} \tag{3.52}$$

which is generally non-zero. This change in energy clearly is reflected in the change in the ADM mass of the perturbed eternal black hole solution. This raises the interesting question that in the case of a single-sided black hole, what is the bulk interpretation of

the fact that the total energy *does not* change (3.51) upon acting by the perturbation, equation (3.47)? We will discuss this question in the next subsection, where we consider the bulk properties and effects of the perturbation.

Notice that if in the thermofield case we consider the first order variation of the *modular* Hamiltonian $H_R - H_L$, then it is zero to first order as $\delta H_R = \delta H_L$. This is analogous to the one-sided case, where expectation value of the modular Hamiltonian $H - E_0$ also does not change as we found above in (3.51).

A more direct method to calculate the energy change is to write the operators in terms of spatial Fourier modes

$$\mathcal{O}(t,\Omega) = \frac{1}{(2\pi)^d}\int d\omega \sum_{lm}\left[e^{-i\omega t}Y_{lm}(\Omega)\mathcal{O}_{\omega,lm} + h.c.\right], \tag{3.53}$$

and similarly for the mirror operators (subject to the cutoff ω_*). Here Y_{lm} denote spherical harmonics on $\mathbb{S}^{d-1}$. Since we will be working with s-waves, we will for now drop the angular momentum indices. We define the two-point function $G(\omega)$ by the equation[13]

$$\frac{1}{Z}\mathrm{Tr}(e^{-\beta H}\mathcal{O}_\omega^\dagger \mathcal{O}_{\omega'}) = \delta(\omega-\omega')G(\omega). \tag{3.54}$$

The KMS condition implies

$$G(-\omega) = e^{\beta\omega}G(\omega). \tag{3.55}$$

At large N we expect that we will have similar results for the pure state:

$$\begin{aligned}\langle\Psi_0|\mathcal{O}_\omega^\dagger\mathcal{O}_{\omega'}|\Psi_0\rangle &= G(\omega)\delta(\omega-\omega') + O(1/N),\\ \langle\Psi_0|\mathcal{O}_\omega\mathcal{O}_{\omega'}^\dagger|\Psi_0\rangle &= e^{\beta\omega}G(\omega)\delta(\omega-\omega') + O(1/N).\end{aligned} \tag{3.56}$$

Finally using the definition of the mirror operators (4.12) we can express the two-point functions between mirror Fourier modes, as well as ordinary and mirror Fourier modes in terms of the single function $G(\omega)$. Putting all this together, we can compute the first order change of the energy

$$\begin{aligned}\Delta E_1 &\equiv ig\int dt' f(t')\,\langle\Psi_0|\,\left[H_0, \mathcal{O}(t')\widetilde{\mathcal{O}}(t')\right]\,|\Psi_0\rangle\\ &\propto ig\int dt' f(t')\iint d\omega_1 d\omega_2(\omega_1-\omega_2)\left(e^{-i(\omega_1+\omega_2)t'} - e^{i(\omega_1+\omega_2)t'}\right)e^{\frac{\beta\omega_2}{2}}G(\omega_2)\,\delta(\omega_1-\omega_2)\\ &= 0.\end{aligned} \tag{3.57}$$

Here we have ignored the bound ω_* on the frequencies for the mirror operators, which does not play a role in this calculation. The fact that $\mathcal{O}$ and $\widetilde{\mathcal{O}}$ have opposite commutators with the CFT Hamiltonian plays an important role in making this energy change equal to zero.

3.2.2.2 Shockwaves in One-Sided Black Hole

We would now like to discuss the bulk interpretation of the state-dependent perturbation

$$H \equiv H_0 + gV(t) = H_0 + g\,f(t)\,\mathcal{O}(t)\,\widetilde{\mathcal{O}}(t). \tag{3.58}$$

[13] As discussed in the previous section, here we think of the Fourier modes $\mathcal{O}_\omega$ in a distributional sense.

This perturbation creates shockwaves of infalling matter both in the right and left region, very similar to those in the eternal black hole. These effects are of order $O(N^0)$ and in that sense they correspond to quantum matter, rather than classical configurations of matter. If the leading order metric is normalized to be $O(N^0)$, then the backreaction of this quantum matter on the geometry modifies the metric only at order $O(1/N^2)$. We do not yet have a complete understanding of the backreacted geometry at order $O(1/N^2)$. The reason is that the operators $\widetilde{\mathcal{O}}$ have been gravitationally dressed with respect to the right. Hence the quantum matter that they create in the left exterior region of the Penrose diagram should be accompanied by appropriate gravitational Wilson lines, which extend all the way from the left region toward the CFT on the right. These gravitational Wilson lines have to be taken into account when considering the correction to the metric at the $O(1/N^2)$ order.

In order to compute the $O(N^0)$ modification of the quantum state of the fields we follow a procedure similar to that discussed in subsection 3.2.1. We first compute the quantum-corrected bulk two-point function $G(x,x')$ of a scalar field $\phi(x)$ that is dual to the operator $\mathcal{O}$ used in the double-trace perturbation (3.58). This leads to an equation very similar to (3.41)

$$\begin{aligned} G(x,x') &\equiv \langle\Psi_0|\widehat{\mathcal{U}}_g^\dagger\,\phi(x)\,\phi(x')\,\widehat{\mathcal{U}}_g|\Psi_0\rangle \\ &= G_0(x,x') - ig\bigg(\int_0^t dt_1\, f(t_1)\,\langle\Psi_0|\big[\mathcal{O}(t_1)\,\widetilde{\mathcal{O}}(t_1),\phi(t)\big]\,\phi(t')\,|\Psi_0\rangle \\ &\quad + \int_0^{t'} dt_2\, f(t_2)\,\langle\Psi_0|\,\phi(t)\,\big[\mathcal{O}(t_2)\,\widetilde{\mathcal{O}}(t_2),\phi(t')\big]\,|\Psi_0\rangle\bigg), \end{aligned} \tag{3.59}$$

where $\widehat{\mathcal{U}}_g$ is given in (3.49). From the definition of the mirror operators (4.12), it follows that the correction to the bulk two-point function in the typical state is the same as that in the eternal black hole (3.41). Hence, at order $O(N^0)$, we find that the scalar field $\phi(x)$ has the same bulk stress tensor as the one discussed in subsection 3.2.1. This stress tensor corresponds to a shockwave falling into the black hole from the right region, as can be checked by direct calculation.

A similar calculation can be done for the left region. We consider the part of the left region which is within a few Schwarzchild radii from the bifurcation point. In that region the local bulk field $\widetilde{\phi}$ can be reconstructed, for example by an analogue of the HKLL prescription, where we will use the mirror operators $\widetilde{\mathcal{O}}$ instead of the usual operators $\mathcal{O}$. In that region, and in the limit of large ω_*, the bulk two-points function is the same (up to the obvious left-right reflection) as the bulk two-point function in the right region. We can compute the effect of the perturbation (3.58) by following a similar analysis as in (3.59), with the obvious replacements $\phi \leftrightarrow \widetilde{\phi}$. The final conclusion is that (3.58) produces a shock-wave like stress tensor in the left region.

All in all, we find that to order $O(N^0)$ the perturbation (3.59) creates two shockwaves of infalling matter which are similar to the two-sided case. By selecting the sign of g appropriately we can make sure that these shockwaves have negative null energy. We emphasize that the existence of the shockwave on the right is completely unambiguous as it follows directly from a the algebraic properties of the $\widetilde{\mathcal{O}}$ operators and their effect on HKLL operators via the perturbation (3.58). On the other hand the interpretation of the left shockwave relies on our conjecture about the geometry of the typical state, and that the operators $\widetilde{\mathcal{O}}$ physically describe the left region.

Also notice that while in the figures we depict the left shockwave as if it was coming from a sharply defined region of the left boundary it should be kept in mind that given that we only use frequencies $|\omega| < \omega_*$ the shockwaves are always somewhat smeared in time.

3.2.2.3 Gravitational Wilson Lines and the Backreacted Geometry

In addition to the two shockwaves, the perturbation (3.58) creates gravitational Wilson lines extending all the way to the left. This is because the operators $\widetilde{\mathcal{O}}$ are gravitationally dressed with respect to the right, in particular they do not commute with the CFT Hamiltonian. The gravitational Wilson lines are spherically symmetric. This follows from the definition of the mirror operators (4.12). There we have implicitly assumed that the mirrror operators are defined so that they commute with the boundary stress tensor, once its zero mode has been removed. This means that $[\widetilde{\mathcal{O}}, T'_{00}] = 0$, where $T'_{00} \equiv T_{00} - H/V$, where V is that spatial volume of the sphere where the CFT lives. This is part of the definition of the mirror operators and other choices could be made which would result in non-spherically symmetric gravitational dressings of the mirror operators.

The existence of the gravitational Wilson lines is important in order to understand the vanishing energy change (3.51) at first order in g under the perturbation (3.58). This perturbation creates a negative energy shockwave in the right region. At the same time the perturbation inserts gravitational Wilson lines due to $\widetilde{\mathcal{O}}$, which has positive energy with respect to the CFT Hamiltonian. The Wilson lines encode the CFT energy of the left shockwave. That shockwave has negative local energy, but as it lies in the left region it has positive energy from the point of view of the CFT Hamiltonian, see for example appendix B.3. Considering both effects, and to leading order in g, we find that the energy remains the same and the location of the horizon with respect to the right boundary is unchanged.

If we apply the perturbation (3.58) to a state which contains particles moving in the region behind the horizon, then we need to understand how these excitations are affected by the gravitational Wilson lines. This is equivalent to understanding how to "glue" the geometries, the one before the perturbation (i.e. below the Wilson lines) and the one after the perturbation (above the Wilson lines). This gluing will determine the motion of probes in the geometry. While we have not completed this analysis, the results of the following sections provide evidence that the gluing and the effect of the Wilson lines is such that the trajectories of probes are not significantly affected when crossing the Wilson lines, in the sense that their effect is suppressed at large N. This is to be contrasted with the effect of the right shock-wave on the trajectory of the probe, which is $O(1)$ when the operators are separated by scrambling time.

We notice that the bulk Einstein equations are modified exactly on the Wilson line. This modification refers only to the subleading terms in $1/N^2$. The bulk equations of motion reflect the boundary dynamics. If we consider the time-dependent perturbation (3.58) the boundary equations of motion are modified for a period of time. Hence it is natural that the bulk equations may need to be supplemented by the contribution from the sources.

Before we close this subsection we notice that these subtleties about the effect of gravitational Wilson lines, the question of gluing different geometries along spacelike slices and the modification of the bulk Einstein equations on the gluing surface is not specifically related to the mirror operators, or the conjecture about the geometry of a typical

state. Similar issues arise whenever we consider perturbations of the CFT Hamiltonian by "precursors" and this is generally a topic which deserved further investigation.

For example, suppose we start with the CFT in the ground state $|0\rangle$ on $\mathbb{S}^{d-1} \times$ time. At time $t = 0$ we act with a unitary of the form $U = e^{ig\phi(0)}$ where $\phi(0)$ is an HKLL operator in some particular gravitational gauge. Here the perturbation U is a precursor, which means that while the HKLL operators are usually written as integrals over time, here we use the CFT equations of motion to localize this operator on the boundary at $t = 0$. The question we want to understand is what is the bulk geometry dual to the boundary state, which suddenly switched from $|0\rangle$ for $t < 0$ to $U|0\rangle$ for $t > 0$. We expect that at very early times the bulk geometry should look like empty AdS, while at very late times it will look like AdS with some particles. For intermediate times around $t = 0$ the bulk interpretation is less clear. For instance, suppose we ask what is the backreacted bulk geometry. For $t < 0$ the mass is zero, while for $t > 0$ the mass is nonzero. We need to glue two geometries of different mass. This sudden change of mass is induced by the gravitational Wilson lines. It would be interesting to understand this toy model in more detail. It captures some of the complications that we face when trying to determine the bulk geometry in our case.

These questions are relevant only when we act with precursors, i.e. boundary operators which directly create particles deep in AdS (together with the accompanying gravitational Wilson lines). If we create the particle by switching on the source near the boundary then the geometry can be understood without ambiguity in terms of collapsing matter falling into AdS from infinity.

To summarize, we postpone the interesting question of understanding the bulk geometry to order $O(1/N^2)$ to further work. For now we assume that the boundary arguments presented in the following sections provide evidence that the net effect of probes going through the Wilson lines region is that their trajectory is not drastically modified.

Finally, we mention that for the kind of typical states with narrow energy band that we are considering, it would not be straightforward to gravitationally dress the mirror operators towards the left. This is because the left dressing would require the algebra $[H, \widetilde{\mathcal{O}}] = 0$, which is inconsistent on such states [16].

3.2.3 Probing the Region Behind the Horizon

We have discussed the bulk interpretation of the state-dependent perturbation of the form $\mathcal{O}\widetilde{\mathcal{O}}$. In this subsection we will use this perturbation to probe the different bulk regions and study the horizon.

3.2.3.1 Thought Experiment 1

We will now discuss the first thought experiment. In brief, this experiment is designed to probe regions behind the horizon in the conjectured Penrose diagram 3.1. There are two variants of this experiment, as displayed in figure 3.8. Let us start with the first variant which is displayed on the left in figure 3.8. Here, the orange lines indicate the two shockwaves and the blue line indicates a particle excitation in the left exterior region. In subsection 3.1.5, we argued that such excitations can be obtained by turning on time-dependent sources for the mirror operators in the CFT, say at time $t = -t_*$ where again t_* is of the order of the scrambling time $\frac{\beta}{2\pi} \log S$. Because mirror operators commute with simple operators, such excitations cannot be detected in the CFT using simple operators.

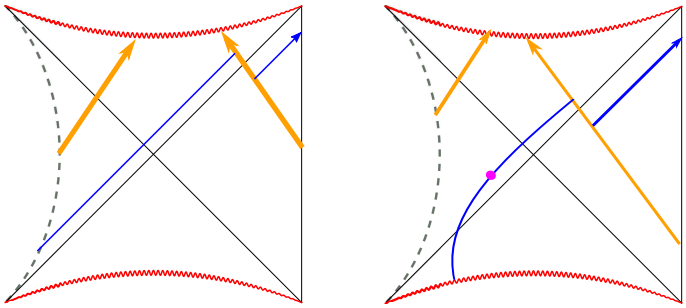


Figure 3.8: Two variants of Experiment 1.

However, if we further perturb the CFT by the state-dependent perturbation V (3.58), the situation changes. We argued in the last Subsection 3.2.2.2 that the state-dependent perturbation produces two negative energy shockwaves on either side of the horizon. The excitation in the left region interacts with the right shockwave and experiences a null shift in the right region. After a finite proper time, it can then be detected in the CFT using a simple operator. We can thus interpret the negative energy shockwaves to have made the horizon traversable.

In order to verify that this indeed happens, we need to compute the following correlator. Turning on the mirror source corresponds to acting with $e^{i\epsilon\widetilde{\phi}(-t_*)}$. Following this, we act with the perturbation at time $t=0$ using the unitary $\widehat{\mathcal{U}}_g = \mathcal{T} e^{ig\int dt' f(t')\,\mathcal{O}(t')\,\widetilde{\mathcal{O}}(t')}$. We then compute the expectation value of $\phi(t)$ on the resulting state. All in all we need

$$\mathcal{C}' \equiv \langle \Psi_0 |\; e^{-i\epsilon\widetilde{\phi}(-t_*)}\, \widehat{\mathcal{U}}_g^\dagger\, \phi(t)\, \widehat{\mathcal{U}}_g\, e^{i\epsilon\widetilde{\phi}(-t_*)}\, |\Psi_0\rangle\,. \tag{3.60}$$

This correlator is analogous to the following correlator (3.37) when we do the same thought experiment in the two-sided black hole

$$\mathcal{C} \equiv \langle \Psi_{\text{TFD}} |\; e^{-i\epsilon\phi_L(-t_*)} (\mathcal{U}_g)^\dagger \phi_R(t)\, \mathcal{U}_g\; e^{i\epsilon\phi_L(-t_*)}\, |\Psi_{\text{TFD}}\rangle\,. \tag{3.61}$$

where $\mathcal{U}_g \equiv \mathcal{T} e^{ig\int dt' f(t') \mathcal{O}_L(t')\,\mathcal{O}_R(t')}$. This was explicitly calculated in [5, 9] and shown to be non-zero. Instead of directly computing $\mathcal{C}'$, we will argue that it is approximately equal to $\mathcal{C}$, which is easier to calculate.

Using the defining equations for the mirror operators 4.12 repeatedly, we can rewrite $\mathcal{C}'$ as the expectation value of a complicated string of *ordinary* (i.e. non-mirror) operators[14] on the state $|\Psi_0\rangle$. We call this string of operators $\mathcal{X}(\phi, \mathcal{O})$, so we have

$$C' = \langle \Psi_0 | \mathcal{X}(\phi, \mathcal{O}) | \Psi_0 \rangle. \tag{3.62}$$

where C' here is the same correlator as the one in (3.60).

Similarly, in the case of the two-sided black hole, the action of $\mathcal{O}_L$ operators can be re-written in terms of the $\mathcal{O}_R$ operators using the properties of the TFD state

$$\begin{aligned}
\mathcal{O}_{L,\omega} |\Psi_{\text{TFD}}\rangle &= e^{-\frac{\beta H_{\text{TFD}}}{2}} \mathcal{O}^\dagger_{R,\omega} e^{\frac{\beta H_{\text{TFD}}}{2}} |\Psi_{\text{TFD}}\rangle, \\
\mathcal{O}_{L,\omega} \mathcal{O}_{R,\omega_1} ... \mathcal{O}_{R,\omega_n} |\Psi_{\text{TFD}}\rangle &= \mathcal{O}_{R,\omega_1} ... \mathcal{O}_{R,\omega_n} \mathcal{O}_{L,\omega} |\Psi_{\text{TFD}}\rangle, \\
[H_{\text{TFD}}, \mathcal{O}_{L,\omega}] \mathcal{O}_{R,\omega_1} ... \mathcal{O}_{R,\omega_n} |\Psi_{\text{TFD}}\rangle &= \omega \mathcal{O}_{L,\omega} \mathcal{O}_{R,\omega_1} ... \mathcal{O}_{R,\omega_n} |\Psi_{\text{TFD}}\rangle.
\end{aligned} \tag{3.63}$$

[14] It is crucial to realize that, after the mirror operators are mapped to normal operators, the resulting correlators do not correspond to experiments that can set up by only using the normal operators.

where $H_{\text{TFD}} \equiv H_R - H_L$. Notice that these equations are completely similar to equations 4.12 if we identify $\mathcal{O}_{L,\omega} \leftrightarrow \widetilde{\mathcal{O}}_\omega, \mathcal{O}_{R,\omega} \leftrightarrow \mathcal{O}_\omega, H_{\text{TFD}} \leftrightarrow H$.

Using the equations (3.63), we can now repeat the same process in correlator $\mathcal{C}$, by replacing $\mathcal{O}_L, \phi_L$ in terms of right CFT operators. In this way we get *exactly the same string* $\mathcal{X}(\phi_R, \mathcal{O}_R)$, now expressed in terms of $\mathcal{O}_R$. This string is a function only of operators in the right CFT, and hence we can compute it by first tracing out the left CFT. Let us drop the R subscript for economy. The correlator $\mathcal{C}$ then becomes a thermal correlator in the right CFT

$$\mathcal{C} = \frac{1}{Z}\text{Tr}\big(e^{-\beta H}\mathcal{X}(\phi, \mathcal{O})\big), \tag{3.64}$$

here $\mathcal{X}(\phi, \mathcal{O})$ is exactly the same string as the one in (3.62). We know that the correlator $\mathcal{C}$ contains a signal corresponding to the probe traversing the horizon. If the correlator $\mathcal{C}'$ is close to $\mathcal{C}$ then the same signal will be present in the one-sided black hole perturbed by the state-dependent operator (3.58), which will be evidence that a particle was extracted from the left region of our conjectured geometry. This brings us to the main conclusion:

> The conjecture that the bulk geometry of a typical state is described by the Penrose diagram discussed in section 3.1 and that it responds to perturbations in the way predicted by effective field theory on this diagram, requires as a necessary condition that the correlators C, C' are the same at large N. This is essential to hold even when the time separations of the operators are taken to be of the order of scrambling time.

Thus, we have identified a technical condition for CFT correlators, necessary for the smoothness of the horizon of a typical state. We discuss this condition in more detail in section 3.4. We also provide some preliminary evidence in favor of its validity.

A variant setup

Now we come to the second variant of experiment 1, depicted in the right part of figure 3.8. Here, we do not use a time-dependent source for the particle in the left region. Instead of starting with the typical state $|\Psi_0\rangle$ and acting with the operator $e^{i\epsilon\widetilde{\phi}(-t_*)}$, we start in the state

$$|\Psi_I\rangle \equiv e^{i\epsilon\widetilde{\phi}(-t_*)}\,|\Psi_0\rangle\,. \tag{3.65}$$

This is an *autonomous* non-equilibrium state, owing to the fact that $[H, \widetilde{\phi}] \neq 0$. As such, this state is not typical under the Haar measure, but it is an autonomous state in the full CFT Hilbert space nonetheless. Detailed discussion of such states can be found in [25]. The experiment then consists of acting on such a non-equilibrium state by the unitary of the state-dependent perturbation. As before, one then aims to detect the mirror excitation inherent to this state by using a simple operator ϕ. The entire experiment can be encoded in the correlator

$$\mathcal{C}'' = \langle\Psi_I|\,\widehat{\mathcal{U}}_g^\dagger\,\phi(t)\,\widehat{\mathcal{U}}_g^\dagger\,|\Psi_I\rangle\,, \tag{3.66}$$

where V is as before given by equation (3.58). The value of this correlator is closely related to that of $\mathcal{C}'$ in (3.60). It can then be compared to similar correlators in the thermofield setup where the left side of the eternal black hole is in some autonomous non-equilibrium state.

3.2.3.2 Thought Experiment 2

We now study a second thought experiment to probe the region behind the horizon, displayed in figure 3.9. In this experiment, we start with the typical pure state $|\Psi_0\rangle$. Then we

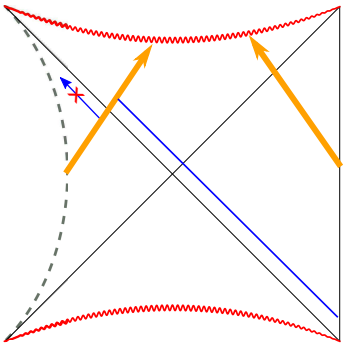


Figure 3.9: Experiment 2

act on this state by a unitary of a simple operator $e^{i\epsilon\phi(-t_*)}$, where t_* denotes a timescale of the order of scrambling time and ϕ a field near the boundary of the CFT. This creates a particle excitation in the right exterior region of the bulk, outside the black hole horizon. We have depicted this using the blue ray in figure 3.9. The CFT state then becomes

$$|\Psi_E\rangle \equiv e^{i\epsilon\phi(-t_*)}\,|\Psi_0\rangle\,. \tag{3.67}$$

The particle depicted in the figure falls towards the black hole and eventually crosses the black hole horizon to go into the interior region. Our goal now is to somehow reconstruct the state of this particle. There are many way to do this, since in principle the information of the particle is always present in the CFT. Here we will describe a protocol which uses the state-dependent perturbations (3.58) and in a particular extrapolation it realizes the Hayden-Preskill protocol as formulated by [9]. This will be the subject of section 3.5.

The protocol is as follows: after throwing the particle in the black hole (3.67) and waiting for scrambling time, we act on the above state at time $t = 0$ by the state-dependent perturbation (3.58). This perturbation creates the shockwaves that we discussed. As shown in the figure 3.9 the shockwaves deflect the particle which moves into the left region. There it can be detected by measuring a mirror operator of the form $\widetilde{\phi}(t_*)$. We thus need to calculate the correlator

$$\mathcal{C}_2' \equiv \langle\Psi_0|\; e^{-i\epsilon\phi(-t_*)}\,\widehat{\mathcal{U}}_g^\dagger\,\widetilde{\phi}(t_*)\,\widehat{\mathcal{U}}_g\,e^{i\epsilon\phi(t_*)}\;|\Psi_0\rangle\,. \tag{3.68}$$

Using similar steps as before we can reduce this correlator to a correlator of ordinary single trace operators and compare it to the corresponding correlator in the TFD state.

Summary

In this section we described some thought experiments, which indirectly probe the region behind the black hole horizon. We showed that our conjecture for the geometry presented in section 3.1 requires as a necessary condition that certain CFT correlators on typical pure states are close to thermal correlators. Assuming that the correlators are indeed the same at large N, we find that the typical black hole microstate responds to perturbations as if it contained the part of the extended Penrose diagram presented in section 3.1.

3.3 The SYK model as an example

We will exemplify some of the previous statements in the context of the SYK model. The SYK model is a toy model of holography, and although it is not expected to have an Einstein bulk dual, it still captures some important features of the bulk theory.

3.3.1 Brief Review of the SYK model

The Sachdev-Ye-Kitaev (SYK) model [41, 42, 43] is a one-dimensional quantum mechanics model containing N species of Majorana fermions ψ_i, $i = 1, 2, ..., N$. The fermions satisfy $\{\psi_i, \psi_j\} = \delta_{ij}$. In general, the fermions in the SYK model have q-body random interactions such that the Hamiltonian is

$$H = (i)^{q/2} \sum_{1 \leq i_1 < i_2 < \cdots < i_q \leq N} J_{i_1 i_2 \cdots i_q} \psi_{i_1} \psi_{i_2} \cdots \psi_{i_q} \,, \tag{3.69}$$

where the coupling constants $J_{i_1 i_2 \cdots i_q}$ are all chosen randomly from a Gaussian distribution with mean zero and variance

$$\langle J^2_{i_1 i_2 \cdots i_q} \rangle = \frac{2^{q-1} (q-1)! \, \mathcal{J}^2}{q \, N^{q-1}} \,. \tag{3.70}$$

The parameter $\mathcal{J}$ has dimensions of energy and sets the scale of the problem. The variance of the coupling $J_{i_1 i_2 \cdots i_q}$ is chosen to depend explicitly on N so that the model has interesting properties in the large N limit. When $q = 2 \bmod 4$, the factor of $(i)^{q/2}$ upfront is necessary to make the Hamiltonian Hermitian. The model becomes conformal at low energies i.e. when the frequencies are very small compared to $\mathcal{J}$. The conformal limit of this model has been studied in detail in [42, 43].

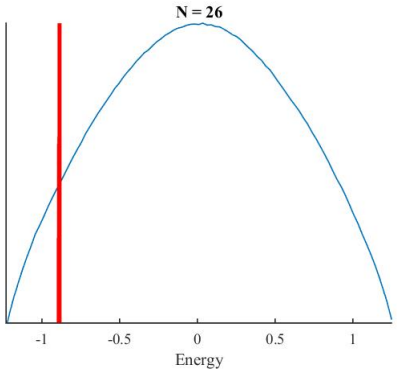


Figure 3.10: Distribution of energy eigenvalues of the SYK model, the red strip shows the energy eigenstates selected for a typical state.

It is easier to compute correlation functions in the SYK model after taking an ensemble average over the couplings $J_{i_1 i_2 \cdots i_q}$. However, we will assume a particular realization of the coupling constant $J_{i_1 i_2 \cdots i_q}$ to obtain a unitary model. This is not a problem because the SYK model is self-averaging: to leading order at large N correlators are the same if we choose a specific realization of the coupling constants $J_{i_1 i_2 \cdots i_q}$ or perform a

disorder-average over it[15]. Thus, at large N we are in principle able to compare correlators calculated in a particular realization (say numerically) with the ones estimated (eg. analytically) using the disorder-average. For finite N, the Hamiltonian is a finite-dimensional matrix with size $2^{N/2} \times 2^{N/2}$ and has $2^{N/2}$ energy eigenvalues. It is relatively easy to find these eigenvalues and corresponding eigenstates by direct numerical diagonalization for reasonably large N. In figure 3.10, we display distribution of energy eigenvalues for $N = 26$.

3.3.1.1 Equilibrium and Non-Equilibrium States in the SYK Model

The finite size of this model (at finite N) makes the SYK-model a good tool to numerically test various statements about typical state and the perturbations discussed earlier in section 3.1.5. We, moreover, have greater analytic control over some aspects of the model, which allows as to do more explicit CFT calculations. Sometimes it is easier to consider a set of spin operators

$$S_k \equiv 2i\psi_{2k-1}\psi_{2k} \quad , \qquad k = 1, \ldots, N/2, \tag{3.71}$$

to further simplify calculations. These operators are bosonic and are therefore more in line with earlier discussions. We will assume that we have a particular realization of the SYK ($q = 4$) couplings J_{ijkl}. This means we have a well defined quantum system with a Hilbert space and unitary time-evolution. Nevertheless we will use results from disorder averaging as a mathematical technique, which allows us to estimate certain correlators for the model with a particular realization of J_{ijkl}, as the disorder in the SYK model is self-averaging.

The Hamiltonian has $2^{N/2}$ energy eigenstates $|E_i\rangle$, which can can be found by any diagonalization method at finite N. The interesting critical behavior of SYK takes place at the low-energy regime of this spectrum. We will define typical pure states in the SYK model by writing down pure states of the form

$$|\Psi_0\rangle \; = \sum_{E_i \in (E_0, E_0+\delta E)} c_i |E_i\rangle, \tag{3.72}$$

where $|E_i\rangle$ are the exact SYK eigenstates (for a particular realization), and we select an energy window $(E_0, E_0 + \delta E)$ centered around some energy E_0[16] and with width δE. We assume E_0 is in the low energy regime, where the SYK model is strongly coupled and $E_0 - E_{\text{gs}} \sim aN$ where E_{gs} denotes the ground state energy in SYK model and a is a small number ($a << J$) which does not scale with N. From basic thermodynamics and using the partition function $Z(\beta) = \sum_i e^{-\beta E_i}$ we can relate E_0 to β. We want to be in the regime where $\beta\mathcal{J} \gg 1$. We take the spread δE to scale like $O(N^0)$ which implies that it is very small compared to $E_0 - E_{\text{gs}}$. At the same time we take δE is large enough, so that we have exponentially many (in N) states contributing to equation (3.72).

Let us now consider some examples of exciting an equilibrium state in the SYK model. Usual excited states can be written as

$$e^{i\epsilon S_i(t_0)}|\Psi_0\rangle. \tag{3.73}$$

[15] Notice that self-averaging is expected only for a class of simple correlators and for relatively small time separations (of the order of scrambling time). For long time scales self-averaging may not be true, see for example discussions in [44, 45].

[16] E_0 should not be confused with the energy of the ground state of the SYK model.

The analogue of states with excitations behind the horizon can be written as

$$e^{i\epsilon\widetilde{S}_i(t_0)}|\Psi_0\rangle = e^{-\frac{\beta H}{2}} e^{i\epsilon S_i(t_0)} e^{\frac{\beta H}{2}} |\Psi_0\rangle. \tag{3.74}$$

Adding excitations behind the horizon lowers the energy of the state as shown in (3.28). Hence, in states of the form (3.74) the amplitudes of lower energy eigenstates are amplified relative to $|\Psi_0\rangle$, but coefficients of higher energy eigenstates are also turned on therefore "borrowing" that part of the Hilbert space. This explains why there is no paradox that the fixed, *invertible* operator $e^{-\frac{\beta H}{2}} e^{i\epsilon S_i(t_0)} e^{\frac{\beta H}{2}}$ lowers the *expectation value* of the energy of a typical state, as discussed in more detail in [25]. In figure 3.11 we can explicitly see this effect in the case of the SYK model.

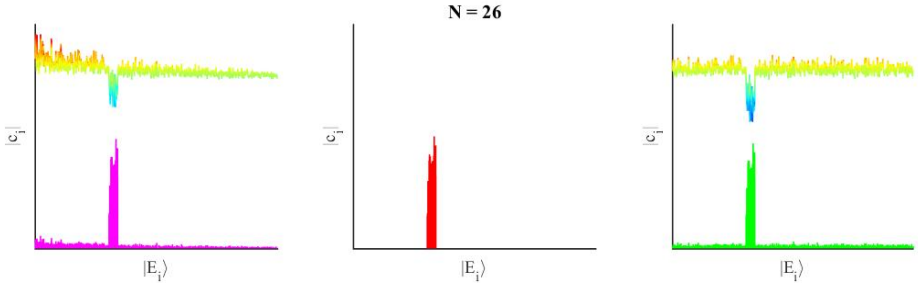


Figure 3.11: Distribution of $|c_i|$ in non-equilibrium state of the form $U(\widetilde{S}_i)|\Psi_0\rangle$ (magenta), typical equilibrium state $|\Psi_0\rangle$ (red), and non-equilibrium state of the form $U(S_i)|\Psi_0\rangle$ (green). The line above the bar plot shows, in heat map colors, which eigenstates are excited, and which ones are suppressed because of the perturbation. Blue eigenstates are suppressed, while eigenstates with other colors are excited with small (green), medium (orange), or large (red) magnitude.

3.3.2 Mirror Operators in the SYK Model

Having defined a typical pure state in the SYK model in equation (3.72), we now define mirror operators. The first step is to define a "small" algebra $\mathcal{A}$ of operators in the SYK model. From the AdS/CFT point of view it would seem more natural to consider only "gauge invariant operators" like $\psi_i\, \partial^{2m+1}\, \psi_i$. The number of such operators at a given conformal dimension does not scale as $N \to \infty$, as expected for CFTs with weakly coupled (but possibly highly curved) AdS bulk duals [46].

However, it is interesting to consider the possibility of defining the mirrors for the non-gauge invariant operators like the individual fermions ψ_i or the spin operators S_k, as many interesting statements about the SYK models can be made directly for the fundamental operators. Thus, we will define the small algebra $\mathcal{A}$ as a span of the low-frequency components of the operators ψ_i and their small products. A typical pure state cannot be annihilated by these operators, hence the construction of the mirror operators can go through. Notice that while the Hamiltonian is quartic in the fermions, it is not part of the algebra $\mathcal{A}$ because we have imposed the condition that only the low frequency components of the fermions are in $\mathcal{A}$, see discussion in subsection 3.1.2, and to reconstruct H we would need the fermions sharply localized at a given moment in time.

We will present the mirror construction for the spin operators S_k, since their bosonic nature makes the presentation simpler. Generalization to the fermions ψ_i is straightforward. The operator $S_k(0)$ can be represented as a $2^{N/2} \times 2^{N/2}$ matrix, by writing the fermions as gamma matrices in the standard basis of Clifford algebra. One can also write it as a matrix in the energy eigenbasis

$$(S_k)_{ij} = \langle E_i | \, S_k \, | E_j \rangle \ . \tag{3.75}$$

Now consider the time evolution of this operator $S_k(t) = e^{iHt} S_k(0) e^{-iHt}$ where H is the SYK Hamiltonian. This defines for us the time-dependent Heisenberg operator. Its exact Fourier modes are

$$(S_k)_\omega^{\text{exact}} \equiv \int_{-\infty}^{\infty} dt \, e^{i\omega t} \, e^{iHt} S_k(0) \, e^{-iHt} \, . \tag{3.76}$$

Using the definition of S_k, this can be written as a product of a sum of delta functions of the form $\delta(\omega - E_i + E_j)$. As we discussed in subsection 3.1.2, to smoothen these delta functions out, we define the coarse-grained Fourier modes, as a function of frequencies in the bin $(\omega, \omega + \delta\omega)$. We take $\delta\omega$ to be of order $K\omega_{\text{gap}}$ where ω_{gap} is the typical gap between energy eigenstates and $K \gg 1$. The coarse-grained modes then become

$$(S_k)_\omega = \frac{1}{\sqrt{\delta\omega}} \int_\omega^{\omega+\delta\omega} d\omega' \, (S_k)_{\omega'}^{\text{exact}} \, . \tag{3.77}$$

An algorithm to find the mirror operators explicitly can be constructed as follows [8]. First we construct a set of vectors which will span the small Hilbert space. We can use the coarse grained spin operators $S_{k,\omega}$ to do this. The set of vectors is constructed as

$$\begin{aligned}
|1\rangle &= |\Psi_0\rangle \, , \\
|2\rangle &= S_{1,\omega_1} |\Psi_0\rangle \, , \\
|3\rangle &= S_{1,\omega_2} |\Psi_0\rangle \, , \\
&\vdots \\
|n\rangle &= S_{2,\omega_1} |\Psi_0\rangle \, , \\
|n+1\rangle &= S_{2,\omega_2} |\Psi_0\rangle \, , \\
&\vdots \\
|l\rangle &= S_{2,\omega_2} S_{1,\omega_1} |\Psi_0\rangle \, , \\
&\vdots
\end{aligned} \tag{3.78}$$

This set of states is not orthogonal but we make sure that we truncate the set so that the states are linearly independent. This truncation is the same as the truncation in the definition of the small algebra $\mathcal{A}$ in subsection 3.1.2. In the large N limit the set of vectors (3.78) can be taken to be very large, and the effects of the truncation become unimportant for our purposes, see [8] for a discussion of the truncation in a more general context.

To simplify the notation, we denote these states as

$$|I\rangle \equiv \mathcal{O}_I |\Psi_0\rangle, \tag{3.79}$$

where $\mathcal{O}_I$ is a combination of the spin operators introduced above. We define $G_{IJ} \equiv \langle I|J\rangle$ as a metric between the states. Since we demanded that the states are linearly independent,

G_{IJ} is an invertible matrix and we denote its inverse by G^{IJ}. We also define

$$B_{IJ,k\omega} \equiv \langle I| \widetilde{S}_{k,\omega} |J\rangle \,, \tag{3.80}$$

which can be computed using the equations for the mirror operators (4.12). We therefore have

$$B_{IJ,k\omega} = \langle \Psi_0 | O_I^\dagger O_J e^{-\frac{\beta H}{2}} S_{k,\omega} e^{\frac{\beta H}{2}} |\Psi_0\rangle. \tag{3.81}$$

This is a matrix element involving only the ordinary operators, hence it may be in principle computed. Finally we can represent the mirror operators explicitly as

$$\widetilde{S}_{k,\omega} = G^{IJ} B_{JK,k\omega} G^{KL} \, |I\rangle \, \langle L| \,. \tag{3.82}$$

While this allows us to explicitly construct the mirror operators, for example in Mathematica, it is much more economic to directly apply the equations from section 3.1.2 to transform the mirror operators to normal operators inside correlators.

3.3.3 Comments on the Kourkoulou-Maldacena States

We would now like to discuss some connections of the results of this chapter to the work of Kourkoulou and Maldacena [24]. That chapter considered a class of pure states in the SYK model, denoted by $|B_s(\beta)\rangle$. To construct these states, we first consider a basis of states $|B_s\rangle$, which are eigenstates of the spin operators

$$S_k \equiv 2i\psi_{2k-1}\psi_{2k} \,, \qquad k = 1, \ldots, N/2 \,. \tag{3.83}$$

These states are defined such that

$$S_k \; |B_s\rangle = s_k \, |B_s\rangle \;, \qquad s = 1, 2, \cdots, 2^{N/2} \,, \tag{3.84}$$

where s denotes the collection $s_k = \pm 1$. The states $\{|B_s\rangle\}$ form a complete basis of states [47], each with mean energy around zero. Evolving the states $|B_s\rangle$ in Euclidean time by

$$|B_s(\beta)\rangle \equiv e^{-\frac{\beta}{2}H} \; |B_s\rangle \;, \tag{3.85}$$

we obtain states with an energy more comparable with the typical state of energy corresponding to inverse temperature β. This defines the states $|B_s(\beta)\rangle$ introduced in [24]. Notice that as written in (3.85) the states are not yet unit-normalized.

The ensemble of states $\{|B_s(\beta)\rangle\}$ has the special property that it is equivalent to the thermal ensemble, in the sense

$$\sum_{s=1}^{2^{N/2}} \langle B_s(\beta)| \, \psi \cdots \psi \, |B_s(\beta)\rangle = \mathrm{Tr}\big[e^{-\beta H} \psi \cdots \psi\big]. \tag{3.86}$$

In the large N limit the SYK model has an approximate $O(N)$ symmetry. A subgroup of this symmetry, identified as the "flip group" in [24], implies that certain flip invariant correlators (for example diagonal two-point functions of the fermions) are the same at large N on all $|B_s(\beta)\rangle$ states, and from (3.86), equal to the thermal correlators. On the other hand there are non flip invariant correlators (for example non-diagonal correlators) which are different from the thermal correlators. In particular they are time-dependent,

indicating that the states $|B_s(\beta)\rangle$ are non-equilibrium states, and in particular *atypical* states. Under time evolution the states thermalize and they equilibrate at late times.

While the states $|B_s(\beta)\rangle$ are atypical, and hence rather different from the typical states which are the focus of this chapter, we will argue that the states $|B_s(\beta)\rangle$ can be approximately constructed by the following procedure. We start with a typical state of energy E_0 in the SYK model and we perform a measurement of many mirror operators $\widetilde{S}_k$. This produces a state which resembles $|B_s(\beta)\rangle$, with a parameter β related to E_0 by the thermodynamics of the SYK model.

To see that we first consider what happens if we perform a measurement of a single mirror-spin operator. It is easier to phrase the discussion in position space, so we introduce the (smeared) position space mirror spins[17]

$$\widetilde{S}_k(t) = \int_{-\omega_*}^{\omega_*} \widetilde{S}_{k,\omega} e^{-i\omega t}. \tag{3.87}$$

As we discussed in section 3.1 these should not be thought of as operators sharply localized in time, but in the limit where ω_* is taken to be large they do start to behave like local operators. In that limit the eigenvalues of $\widetilde{S}_k(t)$ are approximately ± 1. If we measure $\widetilde{S}_k(0)$ we will get either $s_k = 1$ or $s_k = -1$, and the state will be projected to

$$\frac{1 + s_k \widetilde{S}_k(0)}{2} |\Psi_0\rangle. \tag{3.88}$$

Using the definition of the mirror operators we can rewrite this as

$$\begin{aligned} e^{-\frac{\beta H}{2}} \left(\frac{1 + s_k S_k(0)}{2}\right) e^{\frac{\beta H}{2}} |\Psi_0\rangle &= e^{-\frac{\beta H}{2}} \left(\frac{1 + s_k S_k(0)}{2}\right) e^{\frac{\beta H}{2}} \sum_{s'} c_{s'} |B_{s'}(\beta)\rangle , \\ &= e^{-\frac{\beta H}{2}} \left(\frac{1 + s_k S_k(0)}{2}\right) \sum_{s'} c_{s'} |B_{s'}\rangle , \\ &= \sum_{s'} c_{s'} |B_\alpha(\beta)\rangle \, \delta_{s'_k, s_k}. \end{aligned} \tag{3.89}$$

where the last delta function restricts the sum over s' to strings where the k-th spin is required to be s_k.

Now, if we simultaneously measure n mirror-spin operators, and get eigenvalues $s_{i_1}, ..., s_{i_n}$, we will project the state to

$$\left(\frac{1 + s_{i_1} \widetilde{S}_{i_1}(0)}{2}\right) ... \left(\frac{1 + s_{i_n} \widetilde{S}_{i_n}(0)}{2}\right) |\Psi_0\rangle, \tag{3.90}$$

which, using the definition of the mirror operators, can also be written as

$$\sum_{s'} c_{s'} |B_{s'}(\beta)\rangle \, \delta_{s'_{i_1}, s_{i_1}} ... \delta_{s'_{i_n}, s_{i_n}}$$

We notice that as we increase n we fix more and more spins, and if we could extrapolate to $n = \frac{N}{2}$ we would get precisely the $|B_s(\beta)\rangle$ states of [24] as the result of measuring all mirror-spins on a typical pure state.

[17] As discussed before, we may need to add small $1/N$ corrections to make sure that the operators are Hermitian.

This is not an exact statement, for two reasons. First, in defining the mirror operators we had to introduce the cutoff ω_* in the frequencies. Second, we would obtain a single thermal spin state if we projected for all $N/2$ spins, however, this would likely go beyond the small algebra and the mirror operator construction becomes difficult to control before this point.

In [24], Kourkoulou and Maldacena considered state-dependent perturbations of the Hamiltonian of the form $\delta H = g\sum s_k S_k$ on the state $|B_s(\beta)\rangle$, where it was shown that these perturbations lead to an extension of the time evolution from the Rindler patch of AdS_2 to the Poincare patch, thereby gaining access to information from behind the horizon. This effect is present even when we sum over a limited number of spins $K < N/2$, as was shown in [33]. From the arguments above it follows that in such states it is equivalent to write the perturbation as

$$\delta H = g\sum_k S_k \widetilde{S}_k, \tag{3.91}$$

which highlights some similarity with the state-dependent perturbations of the form $\mathcal{O}\widetilde{\mathcal{O}}$ that we have been considering in this chapter. The Kourkoulou-Maldacena perturbation by $\delta H = g\sum_k s_k S_k$ on the state $|B_s(\beta)\rangle$ can be thought as the "quantum-teleportation-version" of the $\delta H = g\sum_k S_k\widetilde{S}_k$ perturbation on a typical pure state $|\Psi_0\rangle$: first we measure $\widetilde{S}_k$ which, as we argued above, transforms the state into $|B_s(\beta)\rangle$. Then, on this state we apply a unitary which depends on the results of the measurement and corresponds to the perturbation $\delta H = g\sum_k s_k S_k$.

3.3.4 Information Behind the Horizon in SYK

In subsection 3.2.3, we discussed two thought experiments to probe the geometry dual to a typical state in a holographic CFT, especially the geometry hidden behind the horizon. We will now illustrate these experiments in the SYK model. Because the details of the two experiments are very similar, we will discuss only the first experiment 3.2.3.1.

We want to consider the analogue of the correlator (3.60) in the SYK model. Hence the correlator we will consider is

$$\mathcal{C}' \equiv \langle\Psi_0|\; e^{-i\epsilon\widetilde{S}_1(-t_*)}\, e^{-igS_2\widetilde{S}_2}\, S_1(t)\, e^{igS_2\widetilde{S}_2}\, e^{i\epsilon\widetilde{S}_1(-t_*)}\,|\Psi_0\rangle\ , \tag{3.92}$$

where $S_2\widetilde{S}_2$ acts $t = 0$, but we suppress this time label for convenience and t_* is scrambling time. We are interested in the term linear in ϵ which is

$$\begin{aligned}\mathcal{C}' &= i\epsilon\,\langle\Psi_0|\left[e^{-igS_2\widetilde{S}_2}\, S_1(t)\, e^{igS_2\widetilde{S}_2}, \widetilde{S}_1(-t_*)\right]|\Psi_0\rangle \\ &= -2\epsilon\,\mathrm{Im}\Big[\,\langle\Psi_0|\, e^{-igS_2\widetilde{S}_2}\, S_1(t)\, e^{igS_2\widetilde{S}_2}\cdot\widetilde{S}_1(-t_*)\,|\Psi_0\rangle\,\Big],\end{aligned} \tag{3.93}$$

where we have rewritten the commutator as the imaginary part of the correlator. The action of the rightmost mirror operator $\widetilde{S}_1(-t_*)$ on the state $|\Psi_0\rangle$ can be replaced by that of $S_1(t_* + i\beta/2)$. Further, the exponential operator can be shown to be

$$e^{igS_2\widetilde{S}_2} = \cos(g)I + i\sin(g)S_2\widetilde{S}_2. \tag{3.94}$$

Using this[18] and the defining equations of the mirror operators the correlator $\mathcal{C}'$ becomes

$$\begin{aligned}\mathcal{C}' = &-2\epsilon\cos^2(g)\,\mathrm{Im}\big[\,\langle\Psi_0|\,S_1(t)\,S_1(t_*+i\beta/2)\,|\Psi_0\rangle\,\big]\\ &-2\epsilon\cos(g)\sin(g)\,\mathrm{Im}\Big[i\,\langle\Psi_0|\,\big[S_1(t),S_2(0)\big]\,S_1(t_*+i\beta/2)\,S_2(i\beta/2)\,|\Psi_0\rangle\,\Big]\\ &-2\epsilon\sin^2(g)\,\mathrm{Im}\Big[\,\langle\Psi_0|\,S_2(0)S_1(t)S_2(0)\,S_1(t_*+i\beta/2)\,\,|\Psi_0\rangle\,\Big].\end{aligned} \tag{3.95}$$

To calculate the first term in equation (3.95) one starts by writing the spin operators in terms of the fundamental fermions $S_j = 2i\psi_{2j-1}\psi_{2j}$. The two-point function in equation (3.95) then becomes a four-point function of four different fermions. In the large N limit of interest, the two-point function is then approximately equal to product of two thermal two-point functions of the fermions. This function can be shown to be real and hence does not contribute to the correlator $\mathcal{C}'$.

We now focus on calculating the second line in equation (3.95). We will first do the calculation in a thermal correlator[19] and later discuss how this approximations the computation in a typical state. So we are interested in the real part of

$$\mathcal{C}' \sim \langle\big[S_1(t),S_2(0)\big]\,S_1(t_*+i\beta/2)\,S_2(i\beta/2)\rangle_\beta\,. \tag{3.96}$$

we will not keep track of overall $O(N^0)$ real multiplicative constants as we are interested in whether there is a signal or not. We will keep track of the phase for the correlators as we need the real part of this correlator. This correlator is a fermion eight-point function

$$\mathcal{C}' \sim \Big\langle\big[\psi_1(\tau_1)\psi_2(\tau_2),\psi_3(\tau_3)\psi_4(\tau_4)\big]\,\psi_1(\tau_5)\psi_2(\tau_6)\,\psi_3(\tau_7)\psi_4(\tau_8)\Big\rangle_\beta\,, \tag{3.97}$$

where the time arguments are determined by equation 3.96.

The easiest way to calculate the leading term is to do this calculation in the vacuum state and then apply a conformal transformation to the thermal circle and analytically continue to Lorentzian time. The Feynman graphs factorize in several pieces at different orders of N [48].

Products of four two-point functions.	$O(N^0)$
Products of a connected four-point function and two two-point functions.	$O(N^{-1})$
Products of two connected four-point function.	$O(N^{-2})$
Products of a connected six-point function and a two-point function.	$O(N^{-2})$
Connected eight-point functions.	$O(N^{-3})$

There are more terms in general, however, terms containing an odd number of fermions are suppressed in the SYK model. The commutator further reduces this list as in some cases the order does not matter, for example in the case of the two-point functions. We, therefore, focus on the instances where the ordering does matter, such as

$$F = \langle\psi_1(t_*+\Delta t)\psi_3(0)\psi_1(t_*+i\beta/2)\psi_3(i\beta/2)\rangle_{\beta,\text{connected}}, \tag{3.98}$$

[18] The operator $\widetilde{S}_2$ in equation (3.94) commutes with $S_1(t_*+i\beta/2)$. The easiest way to see this is to go to frequency space and observe that the imaginary argument of S_1 just gives a multiplicative factor, a c-number.

[19] This calculation is similar to a traversable wormhole calculation in the thermofield double state.

where we have defined $\Delta t = t - t_*$ for later convenience. Observe that this is an out-of-time-order correlator. A special property of out-of-time-order correlators is that they exhibit exponential growth, a hallmark of quantum chaos. These terms can, therefore, become order $O(N^0)$ with times of the order of the scrambling time.

The six- and eight-point functions will also have exponential growth. We can group operators that are close together in time and next to each other, which results in correlators of the form$\langle \mathcal{O}(\text{late})\mathcal{O}(\text{early})\mathcal{O}(\text{late})\mathcal{O}(\text{early})\rangle$. The time ordering of such a correlator has only one exponential growing factor, which is not enough to compensate the suppression in N at the same time that the four point function becomes $O(1)$. The product of two connected four point functions, on the other hand, has two exponential growth factors, enough to compensate the $1/N^2$ suppression of these terms. Connected SYK four point functions have been calculated in [43]. They take the form

$$\langle \psi_1(t_* + \Delta t)\psi_3(0)\psi_1(t_* + i\beta/2)\psi_3(i\beta/2)\rangle_{\text{chaos}} \sim \frac{i}{N}\frac{e^{\frac{2\pi}{\beta}t_*}}{1+e^{-\frac{2\pi}{\beta}\Delta t}} G(\Delta t - i\beta/2), \tag{3.99}$$

Here $G(t)$ is the fermion two-point function. Therefore, only the square of this four point function contributes as we need the real part of equation (3.96). Thus we obtain the following result for equation (3.92)

$$\mathcal{C}' \sim \frac{1}{N^2} e^{2\frac{2\pi}{\beta}t_*} \left(\frac{e^{\frac{2\pi}{\beta}\Delta t}}{1+e^{\frac{2\pi}{\beta}\Delta t}}\right)^2 G(\Delta t - i\beta/2)^2 \,. \tag{3.100}$$

We can see the signal provided that it was sent at an early time of the order of the scrambling time.

The same steps can be taken for the third line in equation (3.95), however, the result is small compared to this one.

Although this correlator was calculated in the thermal state, we expect that at large N we will have the same result for the typical state. We will discuss the general reasons for this in section 3.4. Numerical evidence for this statement is discussed in subsection 3.3.5. We can also argue that this is true as follows. We first write the typical state as a superposition of the spin states, as discussed in section 3.3.3

$$|\Psi_0\rangle = \sum_s c_s \, |B_s(\beta)\rangle \ . \tag{3.101}$$

Using this expansion for the bra and ket in the correlators that we want to compute, we get diagonal and off-diagonal terms with respect to s. For the diagonal terms, we have correlators in the thermal spin states which are indeed close to thermal correlators [24]. Notice that here we assume that the flip symmetry remains a good approximation even for time scales of order of scrambling time, and in particular the leading large N result for flip invariant correlators is the same in the $|B_s(\beta)\rangle$ states as the thermal states even at scrambling time.

The off-diagonal terms between different thermal spin states constitute the major difference between the typical state vs the thermal correlator. Such a cross term is of the form

$$\langle B_{s'}(\beta/2)| \, A \, |B_s(\beta/2)\rangle \,, \tag{3.102}$$

where A is a combination of fermions given by equation (3.95). It is important to note that each fermion appears an even number of times in this combination. We can relate

one spin state to another by flipping a spin, for example $|B_{s'}\rangle \sim \psi_{2k-1}|B_s\rangle$, this flips the k'th spin. We can use this to rewrite the cross term as

$$\langle B_s(\beta/2)| e^{\beta H/2}\Gamma e^{-\beta H/2} A |B_s(\beta/2)\rangle\,, \tag{3.103}$$

where Γ is the combination of fermions needed to flip the string of spins s' into the string s. In Γ we only use the fermions with an odd label, and each of the used fermion is only used once. Therefore, the combination $e^{\beta H/2}\Gamma e^{-\beta H/2}A$ contains some fermions which appear an odd number of times and correlators of this form are suppressed in the SYK model[20]. The number of cross terms cannot compensate for this suppression as they come with random phases. The sum over the cross terms is, therefore, suppressed compared to the diagonal terms, and we are justified in using the thermal correlators as an approximation for the typical state.

The results are consistent with the calculations done in [9], including the effect that the correlators becomes non-zero directly after the perturbation (though only at $O(1/N)$). This is closely related to what would be "stringy" corrections in other theories.

3.3.5 Numerical Comparison

The correlators (3.92) relevant for our thought experiment, can also be calculated numerically, with the results shown in figure 3.12. It is, unfortunately, not possible to directly compare to the analytic results, as for the values of N that we can practically analyze numerically the behaviors usually associated with early time, exponential growth, and saturation overlap. It is, however, useful to check whether the the results for the typical state and the thermal state are close, and indeed they are. More about the numerics in the SYK model can be found in appendix B.4.

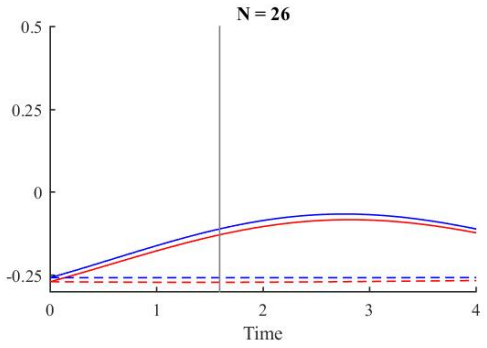


Figure 3.12: Comparing the traversable wormhole correlator (3.92) in the thermal state (blue) and typical state (red). The dotted line is the signal without the probe, which is the response one obtains from just the double-trace perturbation and should disappear in the large N limit, in this case a sum over five pairs of operators was used in the perturbation to limit this effect. The vertical line denotes the scrambling time and is the time around which the probe is focused.

[20] The spin states do have non-zero correlators with an odd number of specific fermions, however, in this case the fermions must pair up for one of the spin operators, which is not the case for the cross terms.

3.4 A conjecture about quantum chaos

In this chapter we made a proposal for the interior geometry of a typical black hole microstate. In section 3.2 we related this proposed interior geometry, to a *necessary* condition for CFT correlators:

> **Conjecture:** In a holographic large N CFT, correlators[21] in typical pure states are close to thermal correlators, even if the time separations $|t_i - t_j|$ are of the order of scrambling time $\beta \log S$.

$$\langle\Psi|\mathcal{O}_1(t_1,x_1)...\mathcal{O}_n(t_n,x_n)|\Psi\rangle_{\text{low-pass}} = \frac{1}{Z}\text{Tr}[e^{-\beta H}\mathcal{O}_1(t_1,x_1)...\mathcal{O}_n(t_n,x_n)]_{\text{low-pass}} + \text{small} \quad \text{error}, \tag{3.104}$$

where "small error" goes to zero as $N \to \infty$. Moreover, the correlators remain close to each other even after analytic continuation to imaginary time, within a strip of at most $t_E = \pm\beta/2$. The subscript "low-pass" means that we only keep the low frequency components $|\omega| < \omega_*$, where ω_* is kept fixed as $N \to \infty$.

A very similar conjecture was formulated earlier in [45], mainly in the context of the SYK model.

Strictly speaking we are interested in this conjecture for holographic CFTs with Einstein gravity duals, however the conjecture may apply to more general theories. For example, we have some partial evidence that it holds in the SYK model whose gravitational dual is not precisely geometric [46].

In the rest of this section, we first make some general comments about the conjecture (3.104) and discuss why it is non-trivial. Then we explain that the conjecture can be simplified by replacing the typical pure state by the microcanonical density matrix. We also discuss some general aspects of comparing canonical and microcanonical ensembles in statistical mechanics. Finally we provide some evidence supporting (3.104). The conjecture (3.104) can be considered as a purely-CFT conjecture, which is independent of the discussion about the bulk interpretation and the interior of a typical black hole microstate.

3.4.1 General Comments on the Conjecture

The subscript "low-pass" in (3.104) means that we only keep the Fourier modes in the correlator with $|\omega| < \omega_*$, where ω_* is a large frequency that we keep fixed as $N \to \infty$. For example, if we have a function of a single variable, we define the low-pass filtered combination by

$$[f]_{\text{low-pass}}(t) \equiv \frac{1}{2\pi}\int_{-\omega^*}^{\omega^*} e^{-i\omega t}d\omega \int_{-\infty}^{+\infty} dt' e^{i\omega t'} f(t') = \frac{1}{\pi}\int_{-\infty}^{+\infty} dt' \frac{\sin[\omega_*(t-t')]}{t-t'} f(t'). \tag{3.105}$$

We can similarly define the low-pass filtered correlators depending on more time-arguments. The low-pass filtering is motivated by the fact that, as discussed in sections 3.1 and 3.2, for the purpose of probing the black hole interior we do not need high frequencies. Moreover, the restriction of frequencies allows us to avoid certain technical problems discussed below.

First, in (3.104) we have not been precise about the nature of convergence between the two correlation functions. For example, even if the correlators converge to each other point-wise in the coordinates, their derivatives may not converge. For instance, suppose

[21] Here we are talking about Wightman correlators, in particular they need not be time-ordered.

that the two correlators differ by a "noise term" with very small amplitude but very high frequency of the form

$$\frac{1}{N}e^{-iNt}. \tag{3.106}$$

Here t denotes schematically some combination of the times in (3.104). While point-wise in t this noise term goes to zero as $N \to \infty$, we notice that if we compute the time-derivatives of the correlators, then they will generally diverge as $N \to \infty$.

Second, a related aspect of this problem is that upon analytic continuation to imaginary time, small differences can get amplified. Consider for example the possible high-frequency noise term (3.106). Typically we would like to compare correlators on a strip of Euclidean width of order β. We notice that upon analytic continuation $t \to t + i\epsilon$, the term blows up exponentially for fixed $\epsilon > 0$ and $N \to \infty$.

Both of these problems are avoided by considering the low-pass-filtered correlators, where only frequencies $|\omega| < \omega_*$ are kept. For example, we avoid problematic terms such as the high frequency, small amplitude noise term (3.106). For the same reason, upon analytic continuation a noise term can be amplified at most by a factor of $e^{\beta\omega_*}$. Hence if the amplitude of the noise term in real time goes to zero as $N \to \infty$, then the same will be true in the complexified time-domain that we are interested in.

We now discuss another aspect of the conjecture (3.104): one may think that the proximity of correlators between typical pure states and the thermal state is an obvious result in statistical mechanics. However condition (3.104) is non-trivial for the following reasons:
(1) There is no general proof about the proximity of all expectation values in typical pure states and the thermal ensemble. The proximity can be established only if some additional assumptions are made about the nature of the observable that we are considering. The Eigenstate Thermalization Hypothesis (ETH) [36] is an example of such an assumption. The observables in (3.104) involve products of operators at large time separations, so it is not automatically obvious that they obey these ETH-like assumptions.
(2) Relatedly, the general expectation in statistical mechanics is that correlators on typical pure states are close to thermal correlators up to $1/S$ corrections. When considering out of time order correlators in large N gauge theories, certain terms which begin at early times as being of order $\frac{1}{N}$, grow as we increase the time separation and at scrambling time they become $O(N^0)$ and mix with leading terms. The conjecture (3.104) states that this effect is the same in the typical pure state and the canonical ensemble.
(3) Condition (3.104) requires that, as we take $N \to \infty$, the observable whose expectation value we are considering also changes, because the scrambling time explicitly depends on N.

3.4.1.1 Replacing Typical Pure States by Microcanonical Mixed State

A second observation is that on very general grounds it can be shown that correlators on typical pure states are very close to those in the micrononical ensemble ρ_m, centered around the appropriate energy window [49]. In particular

$$\langle\Psi|A|\Psi\rangle = \mathrm{Tr}[\rho_m A] + O(e^{-S}). \tag{3.107}$$

Contrary to the approximation between typical pure state and the thermal ensemble, the approximation between a typical pure state and the microcanonical ensemble is much more

robust and we do not expect it to break-down even at late times. The reason is that we can derive rigorous bounds on the variance of expectation values of observables among different pure states, and these bounds depend only on the norm of the observable. For example, it is easy to show [49] that for *any* Hermitian observable A we have

$$\begin{aligned}\int [d\mu](\langle\Psi|A|\Psi\rangle - \ \mathrm{Tr}[\rho_m A])^2 &= \frac{1}{\mathcal{N}+1}\left(\ \mathrm{Tr}[\rho_m A P_m A] - \ \mathrm{Tr}[\rho_m A]^2\right), \\ &\leq \frac{1}{\mathcal{N}+1}\left(\ \mathrm{Tr}[\rho_m A^2] - \ \mathrm{Tr}[\rho_m A]^2\right),\end{aligned} \tag{3.108}$$

where $[d\mu]$ is the usual Haar measure over pure states in the relevant energy window and $\mathcal{N} \sim e^S$ denotes the dimensionality of that Hilbert space. P_m is the projector on that window and $\rho_m = \frac{P_m}{\mathcal{N}}$. The variance on the RHS is exponentially suppressed in S, times a combination of the norms of the operators A and A^2. For us $A = \mathcal{O}(t_1, x_1)...\mathcal{O}(t_n, x_n)$. If we work with local operators which have been smeared in such a way that they are bounded, then separating the constituent operators $\mathcal{O}$ in time cannot increase the norm of the product A, even if the time separation is very large. This follows from two obvious observations. First since time evolution is unitary we have that for any bounded operator

$$||e^{iHt}\mathcal{O}e^{-iHt}|| = ||\mathcal{O}||. \tag{3.109}$$

Second, for any product of bounded operators we have

$$||\mathcal{O}(t_1, x_1)...\mathcal{O}(t_n, x_n)|| \leq ||\mathcal{O}(t_1, x_1)||...||\mathcal{O}(t_n, x_n)||. \tag{3.110}$$

This means that if we work with bounded operators the deviation between typical pure state and microcanonical ensemble cannot grow as a function of time, and hence we can show that

$$\langle\Psi|\mathcal{O}_1(t_1, x_1)...\mathcal{O}_n(t_n, x_n)|\Psi\rangle_{\text{low-pass}} = \mathrm{Tr}[\rho_m \mathcal{O}_1(t_1, x_1)...\mathcal{O}_n(t_n, x_n)]_{\text{low-pass}} + \text{small} \quad \text{error} \tag{3.111}$$

for *all* time separations. Then, using the property (3.111) we can simplify the original conjecture (3.104) to the following equivalent and simplified form

> **Simplified form of the Conjecture:** In a holographic large N CFT correlators in the microcanonical mixed state of energy $E \approx -\partial_\beta \log Z$ are close to thermal correlators, even if the time separations $|t_i - t_j|$ are of the order of scrambling time $\beta \log S$.

$$\mathrm{Tr}[\rho_m \mathcal{O}_1(t_1, x_1)...\mathcal{O}_n(t_n, x_n)]_{\text{low-pass}} = \frac{1}{Z}\mathrm{Tr}[e^{-\beta H}\mathcal{O}_1(t_1, x_1)...\mathcal{O}_n(t_n, x_n)]_{\text{low-pass}} + \text{small} \quad \text{error}, \tag{3.112}$$

where "small error" goes to zero as $N \to \infty$. Moreover the correlators remain close to each other after analytic continuation to imaginary time, in an appropriate domain, i.e. within a strip of at most $t_E = \pm\beta/2$. Again we emphasize that the comparison of the two ensembles is not trivial, for the reasons 1-3 mentioned earlier.

Comments

(a) First we make some clarifying remarks about the order of limits in this conjecture. The precise meaning of (3.112) is: we consider a holographic CFT with a central charge

of order N^2. We consider a fixed number n of smeared, bounded operators approximately localized around space-time points $(t_i = a_i + b_i \log N, x_i)$, with a_i, b_i, x_i fixed. Here we take all time arguments to be *real*. We consider a fixed inverse temperature[22] β and the corresponding canonical ensemble $\rho_\beta = \frac{e^{-\beta H}}{Z}$. We consider the microcanonical ensemble defined by the energy window $(E_0 \pm \Delta E)$ where $E_0 = -\partial_\beta \log Z$ and ΔE fixed. Then we take the $N \to \infty$ limit without changing any of the "fixed parameters". Then the claim is that

$$\lim_{N\to\infty} |\ \mathrm{Tr}[\rho_\beta \mathcal{O}_1(t_1,x_1)...\mathcal{O}_n(t_n,x_n)]_{\text{low-pass}} - \ \mathrm{Tr}[\rho_m \mathcal{O}_1(t_1,x_1)...\mathcal{O}_n(t_n,x_n)]_{\text{low-pass}}| = 0, \tag{3.113}$$

where the subscript *low pass* indicates that we remove frequencies $|\omega| > \omega_*$ as in (3.105), and ω_* is kept fixed as $N \to \infty$.

(b) For $b_i = 0$ then this conjecture reduces to the standard approximation of correlators between canonical and microcanonical ensemble. The non-trivial aspect of the conjecture is when $b_i \neq 0$. In that case we want to make the statement that these "chaos-enhanced $1/N$ corrections" are the same in the pure and thermal states. Technically having $b_i \neq 0$ makes a difference since it means that as $N \to \infty$ we also change the observables that we are considering (by moving the operators in time).

(c) We assume that there the operators $\mathcal{O}_i$ have been smeared in an appropriate way so that they are bounded operators. In particular this will also regulate possible light-cone singularities.

(d) The conjecture formulated as (3.112) is perhaps more conservative than it could be. The correlators could be close to each other even for longer time scales. This, however, is not necessary for the purpose of the thought experiments discussed in this chapter. We have also not been precise about the explicit bounds for the error terms. For the purposes of our chapter it is sufficient that the error goes to zero as $N \to \infty$.

(e) Finally, even though the conjecture (3.112) is formulated for real time arguments, the proximity of the low-pass-filtered correlators after analytic continuation to imaginary time within a strip of width β is guaranteed by the fact that the two correlators do not contain frequencies higher than some fixed frequency ω_*. In the context of black hole physics this has been discussed in earlier works, for example [50], where it was suggested that perhaps the analytically continued correlators contain information about the details of the black hole microstate. However, for the purpose of probing the black hole interior using the thought experiments discussed in this chapter, we seem to be sensitive to the analytic continuation of the low-pass filtered correlators and as we conjecture above, those do not vary significantly among different microstates. In special situations where one can use supersymmetry to obtain greater control, like LLM and LM geometries, the distinction between thermal and typical pure states was discussed in [51] and [52] respectively. See [53] for a summary of these analyses and [54] for a discussion of the distinguishability of typical pure states in connection to the fuzzball proposal.

[22] The temperature must be chosen such that the system is in a black hole phase.

3.4.1.2 Comments on Comparing Canonical to Microcanonical Ensembles

In general the proximity of the two ensembles is based on the following intuition. The two density matrices of the ensembles are

$$\rho_\beta = \sum_i \frac{e^{-\beta E_i}}{Z} |E_i\rangle\langle E_i| \qquad , \qquad \rho_m = \sum_{E_i \in (E_0 \pm \Delta E)} \frac{1}{\mathcal{N}} |E_i\rangle\langle E_i|. \tag{3.114}$$

Consider an observable A, which in our case would be of the form $A = \mathcal{O}_1(t_1, x_1)...\mathcal{O}_n(t_n, x_n)$. The expectation value of A in either ensemble receives contributions *only* from the diagonal matrix elements of A in the energy eigenstates

$$f(E_i) = \langle E_i|A|E_i\rangle. \tag{3.115}$$

First of all we will assume, in the spirit of the ETH, that $f(E_i)$ can be approximated by "reasonably smooth" function $f(E)$. We also assume that the discrete set of states can be described by a smooth density of states $\rho(E)$. Then we have

$$\mathrm{Tr}[\rho_\beta A] = \int_0^\infty dE \frac{1}{Z} \rho(E) e^{-\beta E} f(E), \tag{3.116}$$

$$\mathrm{Tr}[\rho_m A] = \int_{E_0 - \Delta E}^{E_0 + \Delta E} dE \frac{1}{\mathcal{N}} \rho(E) f(E). \tag{3.117}$$

where $\mathcal{N} = e^S$. The usual argument in statistical mechanics is that for systems with many degrees of freedom $\rho(E)$ increases fast with energy, while $e^{-\beta E}$ decreases fast. Hence the product $\rho(E)e^{-\beta E}$ is sharply peaked at a given window of energies which depends on the temperature. If we further assume that $f(E)$ is relatively smooth and slowly varying, then by selecting the window $(E_0 \pm \Delta E)$ of the microcanonical to coincide with the window where $\rho(E)e^{-\beta E}$ peaks, we can establish the approximation between canonical and microcanonical.

This leads to a saddle point approximation, which relies on taking the thermodynamic limit. Usually in AdS/CFT this is achieved by taking $N \to \infty$. Moreover, for the saddle point method to work, it is important that $f(E)$ is a slowly varying function of E. For observables corresponding to small products of operators with time-separations of $O(1)$, we find that at large N the function $f(E)$ actually depends on E only via the temperature. This means that for such observables $\frac{df(E)}{dE} \sim O(1/N^2)$ and the saddle point method is reliable, establishing the equivalence of (3.116) and (3.117). The question is whether the same property of slow variation of $f(E)$ is true for observables like those in (3.112) where some of the time separations scale like $\log(N)$.

To emphasize that the slow variation of $f(E)$ is important for the equivalence of the ensembles, let us mention the following argument. When comparing two density matrices ρ_1, ρ_2 it is useful to consider the trace distance

$$D(\rho_1, \rho_2) \equiv \frac{1}{2} \mathrm{Tr}\big(|\rho_1 - \rho_2|\big), \tag{3.118}$$

where $|X|$ is defined as $|X| \equiv \sqrt{X^\dagger X}$. This characterizes how different the two quantum states are. More precisely for any bounded observable A we have

$$|\mathrm{Tr}\big(\rho_1 A\big) - \mathrm{Tr}\big(\rho_2 A\big)| \leq ||A||\, D(\rho_1, \rho_2). \tag{3.119}$$

The trace distance is positive and is bounded from above by 1, and if it is close to 1 then the two density matrices are in principle "fully distinguishable". When comparing the canonical and microcanonical ensembles the trace distance becomes

$$D(\rho_\beta, \rho_m) = \int_0^{E_0-\delta E} dE \frac{\rho(E)e^{-\beta E}}{Z} + \int_{E_0+\delta E}^{\infty} dE \frac{\rho(E)e^{-\beta E}}{Z} + \int_{E_0-\delta E}^{E_0+\delta E} dE \rho(E) \left| \frac{e^{-\beta E}}{Z} - \frac{1}{\mathcal{N}} \right|. \tag{3.120}$$

It is easy to check that when comparing the canonical ensemble to the microcanonical ensemble for relevant systems[23] then no matter how the window $(E_0 \pm \Delta E)$ is selected, the trace distance between the two ensembles is extremely close to 1. This means that there will always exist some bounded observable A with the property that its expectation value is maximally different in the two ensembles. Such an observable would have the property that $f(E)$ would have to be very rapidly varying with E.

All this shows that the conjectures (3.104) and (3.112) are essentially conjectures about the slow variation of diagonal matrix elements $f(E)$ of A, where $A = \mathcal{O}_1(t_1, x_1)...\mathcal{O}_n(t_n, x_n)$, when the time differences $|t_i - t_j|$ are of the order of scrambling time.

3.4.2 Evidence for the Conjecture

We showed that the original conjecture (3.104) can be simplified to the form (3.112). In this subsection we provide some evidence for it. The evidence we provide does not constitute a proof. However, the statements are should be considered heuristic support for the conjecture.

A class of systems where this conjecture could be investigated more precisely might be large c 2d CFTs with a sparse spectrum. In those theories the relevant correlators can be estimated by a conformal block decomposition [55], where it is expected that the effects at scrambling time are related to the Virasoro identity block. It is, however, also important to check that other blocks do not interfere with those effects, see for example [56]. This is an approach which may be worth investigating in future work.

3.4.2.1 Slow Change with respect to Energy

We now discuss an intuitive argument suggesting the slow variation with energy of the diagonal matrix elements of the relevant observables supporting the conjecture formulated in this section. We generally expect that in AdS/CFT thermal correlators of local operators separated by short time scales depend on the energy only via the temperature, hence the diagonal matrix elements obey[24], $\frac{df(E)}{dE} \sim O(1/N^2)$, where we used the notation $f(E_i) = \langle E_i|A|E_i\rangle$. This justifies the use of the saddle point methods in comparing canonical to microcanonical. When we want to separate the operators by scrambling time, then as we take the large N limit we need to *tune* the operator A in an N-dependent fashion, since scrambling time for a fixed temperature depends on N via $\log N$. Hence when we apply the saddle point method at large N we need to take into account that the operator A itself will depend on N, we denote this as A_{t_S}.

The bulk computation in the eternal black hole suggests that the thermal expectation value of A_{t_S} depends on the mass of the black hole only via the temperature, in the sense

[23] For example this can be checked for 2d CFTs where $\rho(E)$ is given by the Cardy formula.

[24] Here we consider —for example— the $\mathcal{N} = 4$ SYM where in the high temperature phase $E \propto N^2R^3T^4$, where R is the size of the sphere on which the CFT lives.

that

$$I \equiv \frac{d}{d\beta} \operatorname{Tr}\left(\rho_\beta A_{t_S}\right) = O(1), \tag{3.121}$$

where we need to differentiate wrt β both the thermal density matrix and the observable which depends on t_S. We can write these two contributions as

$$I = \operatorname{Tr}\left(\frac{d\rho_\beta}{d\beta} A_{t_S}\right) + \operatorname{Tr}\left(\rho_\beta \frac{dA_{t_S}}{dt}\right)\frac{dt_S}{d\beta}. \tag{3.122}$$

This equation is rather schematic as the observable A can have many time arguments of the order of the scrambling time. We expect that the correlator $\operatorname{Tr}[\rho_\beta \frac{dA_{t_S}}{dt}]$ is at most $O(N^0)$, as the growth of correlators is bounded by the Lyaponov exponent [29]. On the other hand we have $\frac{dt_S}{d\beta} = O(\log N)$, hence the second term in (3.122) grows at most like $\log N$. Since the sum of the two terms in (3.122) is $O(N^0)$ we conclude that the first term

$$\operatorname{Tr}\left(\frac{d\rho_\beta}{d\beta} A_{t_S}\right) = O(\log N). \tag{3.123}$$

From this we can estimate the energy dependence of the diagonal matrix elements of the observable A_{t_S}, leading to

$$\frac{df(E)}{dE} \sim O\left(\frac{\log N}{N^2}\right). \tag{3.124}$$

We, therefore, see an enhancement of the error term of the saddle point method, but the error term is still suppressed in N. This justifies the use of the saddle point method and the approximation of the canonical and microcanonical ensemble for observables at scrambling time. This slow variation with respect to temperature can also be seen numerically in figure 3.13.

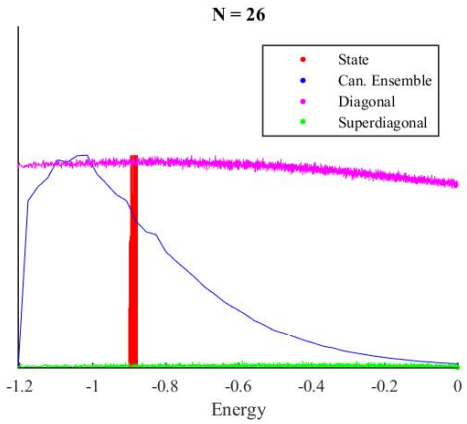


Figure 3.13: The diagonal (magenta) matrix elements of $\{\psi_1(0), \psi(t_*)\}^2$, where t_* is the scrambling time, vary slowly in the SYK model. They, moreover, dominate the elements just above the diagonal (green). The regions important for the typical state (red) and the canonical ensemble (blue, $\rho(E)e^{-\beta E}$) are shown as well. The operators were not passed through a low pass filter.

3.4.2.2 Connection to ETH

The Eigenstate Thermalization Hypotheses (ETH) [36] proposed that the matrix elements of an operator $V_{ij} = \langle E_i | V | E_j \rangle$ have the form

$$V_{ij} = f_V(E)\delta_{ij} + e^{-S(E)/2} g_V(E,\omega) R^V_{ij}, \tag{3.125}$$

where $E_i = E - \omega/2$, $E_j = E + \omega/2$. The functions f_V, g_V are assumed to be slowly varying and R^V_{ij} are some almost random phases. We will discuss to what extent the product of operators obeying the ETH does also obey the ETH. We are interested in timescales of order of scrambling time $t_S = \frac{\beta}{2\pi} \log S$. For concreteness we consider, for example, the correlator $\langle [W(t), V(0)]^2 \rangle$. The matrix elements for $W(t)$ can be written in a similar fashion

$$W(t)_{ij} = f_W(E)\delta_{ij} + e^{-i\omega t} e^{-S(E)/2} g_W(E,\omega) R^W_{ij}. \tag{3.126}$$

To simplify the discussion we can assume that the operators have vanishing one-point functions, i.e. $f_V = f_W = 0$ (or more generally we would have to consider connected correlators). We define

$$C \equiv W(t)V(0), \tag{3.127}$$

which will help us obtain the out-of-time order correlator mentioned above. We have

$$C = \sum_k \Big[e^{-i(E_k - E_i)t}\, e^{-\{S((E_k+E_i)/2) + S((E_k+E_j)/2)\}/2} \times g_W\left(\frac{E_i + E_k}{2}, E_k - E_i\right) g_V\left(\frac{E_j + E_k}{2}, E_j - E_k\right) R^W_{ik} R^V_{kj} \Big]. \tag{3.128}$$

We also need the other ordering, which we will call C'. It has a similar expression with some of the indices interchanged. To obtain an expression for the out-of-time ordered correlator we need terms of the form C^2, C'^2, CC', and $C'C$. The phases are almost random but they are assumed to have some correlation [13]

$$R^V_{ij} R^W_{kl} = h^{VW}(E_i, E_l)\delta_{il}\delta_{kj} + \text{erratic}, \tag{3.129}$$

where $h^{VW}(E_i, E_l)$ is some smooth function. For short time scales this ansatz leads to the conclusion that the combination $[W(t), V(0)]^2$ obeys the ETH, if W and V do.

We will now comment on the opposite regime. For very large time scales, the time-dependent phases in C and C' fluctuate rapidly and the correlations (3.129) are washed out when h^{VW} is inserted in (3.128). In the terms C^2 and C'^2, the matrix product implies that we need to fix two indices (say k, ℓ). As a result we are left with a sum over only one index. This sum is of order $O(e^S)$ but it cannot compensate the $O(e^{-2S})$ suppression. Thus we can neglect these terms. In the cross terms CC' and $C'C$, the phases can be summed over in a coherent way and they then compensate the $O(e^{-2S})$ suppression. We will thus only look at these. If we redefine $E_k = E_i + \omega$ and $E_j = E_i + \omega_1 + \omega_2$, and turn the sums into integrals, we get

$$CC' = \int d\omega_1 d\omega_2 e^{\beta\omega_1/2 + \beta\omega_2/2} |g_W(E_i + \omega_1/2, \omega_1)|^2 |g_V(E_i + \omega_1 + \omega_2/2, \omega_2)|^2, \tag{3.130}$$

where we used that the entropy S varies slowly and that moderate values of ω dominate. ETH assumes that g_V and g_W are slowly varying functions in the first argument. We can, therefore, use a Taylor expansion to obtain

$$CC' + C'C = 2\langle V^2\rangle\langle W^2\rangle + \alpha_V \partial_\beta \langle W^2\rangle + \alpha_W \partial_\beta \langle V^2\rangle, \tag{3.131}$$

$$\alpha_V = 2\int d\omega e^{\beta\omega/2}\partial_E |g_V(E,\omega)|^2. \tag{3.132}$$

Generally speaking one expects that $\alpha_V \ll \langle V^2\rangle$. This expectation can be motivated by analyzing the two-point function, which has the form

$$\langle E|\, V(t)V(0)\,|E\rangle = \int d\omega\, e^{\beta\omega/2 - i\omega t}\left(|g_V(E,\omega)|^2 + \frac{\omega}{2}\partial_E |g_V(E,\omega)|^2\right), \tag{3.133}$$

upon using the ETH ansatz. As discussed in the previous section 3.4.2.1, we expect the derivative of the first term to be very small, thus making α_V small. There is an independent argument to justify that $\alpha_V \ll \langle V^2\rangle$. Notice that it is the connected diagrams which contribute to α_V. These are suppressed in the large N limit as opposed to the disconnected ones that contribute to $\langle V^2\rangle$, leading us to the conclusion $\alpha_V \ll \langle V^2\rangle$. This means that for very long time scales the factorized result dominates, exactly what one expects in the region of saturation after chaotic growth.

The ETH, therefore, works with both short times and very long times. It is natural to expect that ETH will also work for intermediate times, but it seems that we need to make some additional assumptions about the matrix elements of the observables in order to prove this. See [57] for a discussion about possible extensions of ETH to higher order correlators. We postpone further investigation of these questions to future work.

3.4.2.3 Time-Order vs Out-of-Time-Order Correlators

We provide another approach for readers who accept that time-ordered correlators continue to factorize at time scales of the order of the scrambling time. We will argue that this implies the validity of our conjecture also for out-of-time-order correlators.

The exponential growth of parts of chaotic correlators is seen in correlators that are out-of-time-order, for example

$$\langle \mathcal{O}_1(t)\mathcal{O}_2(0)\mathcal{O}_1(t)\mathcal{O}_2(0)\rangle. \tag{3.134}$$

Factorization breaks down in such a case and some $1/N$ corrections are enhanched to $O(1)$ at time scales of the scrambling time. This is specific to out-of-time-order correlators.

The time-order and out-of-time-order correlators stated above are related by the commutator: $[\mathcal{O}_1(t), \mathcal{O}_2(0)]$. These commutators have been studied in the case of 2d CFTs [58], with the assumption that we can promote coordinate transformations $\xi(x)$ to Goldstone fields $\hat{\xi}(x)$ and that we write light operators as functions of these fields. In that case, it was argued that the commutator is dominated by a single term around scrambling time

$$[\mathcal{O}_1(t_1), \mathcal{O}_2(t_2)] \sim \tfrac{1}{c} e^{\lambda(t_1 - t_2)} \partial_{t_1}\mathcal{O}_1(t_1)\partial_{t_2}\mathcal{O}_2(t_2). \tag{3.135}$$

where c is the central charge, and λ is the Lyapunov exponent giving the rate of exponential growth of out-of-time-order correlators. The commutator can be used to transform the out-of-time-order correlator to a time-order correlator, and shows how the exponential growth remains present after we convert the out-of-time-order correlators to time-order correlators.

Nevertheless, the out-of-time-order correlator is the same in the microcanonical and the canonical ensemble if it is dominated by finitely many time-ordered terms that one obtains after using commutators such as those in equation (3.134). In passing, we note that these commutators are not known or easily constructed in general.

3.4.2.4 SYK Numerics

We can numerically check the conjecture and the various statements of this section in the SYK model, see [45] for an in-depth numerical study of the ETH and equivalence between microstate and thermal correlators in SYK. For example, in figure 3.14 we show the agreement between various simple correlators in the canonical ensemble and a typical pure state. The conjecture becomes non-trivial for out-of-time-order correlators with times

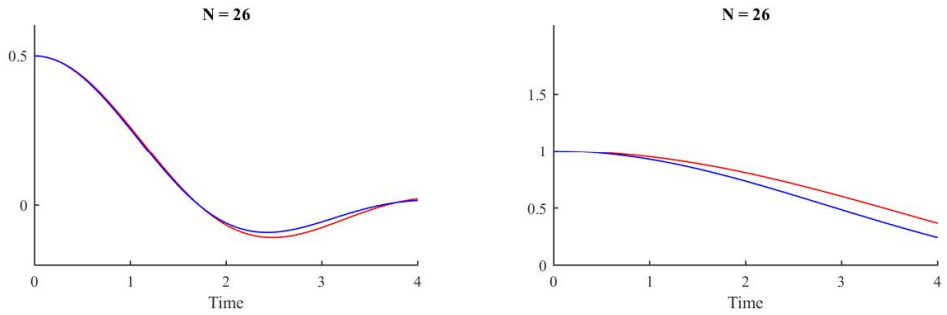


Figure 3.14: We compare the expectation value of $\langle \psi_1(t)\psi_1(0)\rangle$ (left) and $\langle S_1(t)S_1(0)\rangle$ (right) in the thermal state (blue) and pure state (red). We expect the two curves to differ by corrections suppressed by $1/N$. The observed deviation is consistent with the value of N that we used.

of the order of the scrambling times. We, therefore, look at the OTOC in figure 3.15. We can see that the pure and thermal expectation value are very close to each other up to the scrambling time, and close to each other after scrambling time.

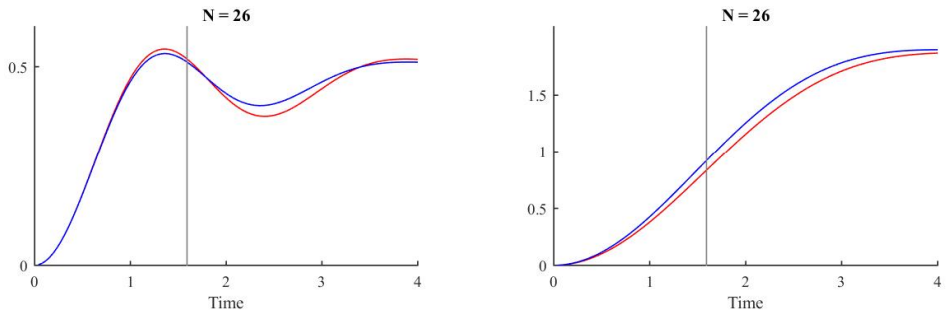


Figure 3.15: We compare the expectation value of $\langle \{\psi_1(t), \psi_2(0)\}^2\rangle$ (left) and $-\langle [S_1(t), S_2(0)]^2\rangle$ (right) in the thermal state (blue) and pure state (red). The scrambling time is designated by the vertical line.

We can also check the arguments about the slowly varying energy dependence of the operators at scrambling time, see figure 3.16. This corresponds to the moment of the grey line in figure 3.15.

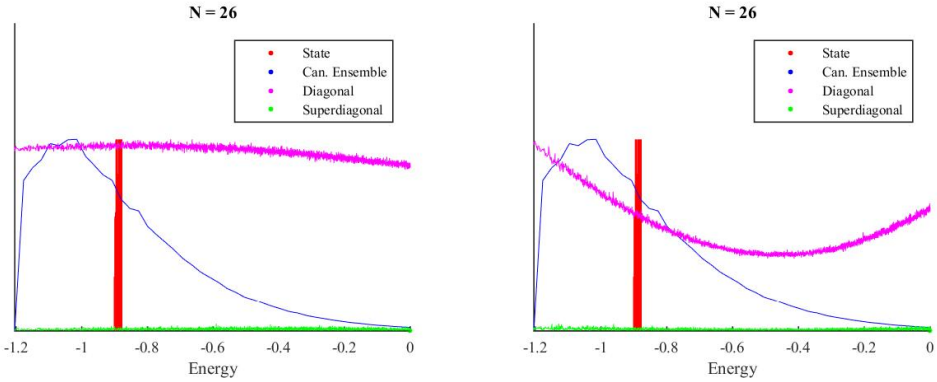


Figure 3.16: We compare the diagonal (magenta) matrix elements of $\{\psi_1(t_*), \psi_2(0)\}^2$ (left) and $-[S_1(t_*), S_2(0)]^2$ (right), where t_* is the scrambling time. They, moreover, dominate the elements just above the diagonal (green). The regions important for the typical state (red) and the canonical ensemble (blue, $\rho(E)e^{-\beta E}$) are shown as well. The operators were not passed through a low pass filter.

These low N calculations seem to support the conjecture. However, the N used in these calculations is too low to clearly see some desired features of the correlators and it would be interesting to numerically study larger values of N. In appendix B.4 we give more details about numerics in the SYK model.

3.5 The mirror operators and the Hayden-Preskill protocol

In this section we discuss how the mirror operators can be used to extract information from a black hole and highlight similarities with the Hayden-Preskill protocol. We start with a review of the original Hayden-Preskill argument and then we discuss the implementation of the protocol suggested by [9]. After that we explain the connection with the mirror operators and the state-dependent perturbations that we have been discussing in this chapter.

3.5.1 The Hayden-Preskill Protocol

In [30], Hayden and Preskill proposed a way to extract information from old black holes i.e. those that have evaporated away at least half of their entropy. Their thought experiment is represented diagrammatically in figure 3.17. Alice sends a message into a black hole after the half-point of evaporation, also called the Page time. This message is drawn as a red ray in figure 3.17. Bob, who wants to spy on Alice's message, has been collecting the Hawking radiation since the moment the black hole started emitting it. He has infinite resources and can study correlations between Hawking quanta exactly. He also knows the exact initial microstate of the black hole and the exact dynamics of quantum gravity governing black hole evaporation, denoted U in the figure. Then the question [30] asked is - how many Hawking particles does Bob need to collect additionally, in order to reconstruct the message that Alice threw into the black hole?

In figure 3.17 time flows upwards. Bob can build a decoder using the early Hawking radiation and part of the late Hawking radiation, represented by the operation $\widetilde{U}$. We assume that the internal dynamics of the black hole U is a random and rapidly-mixing

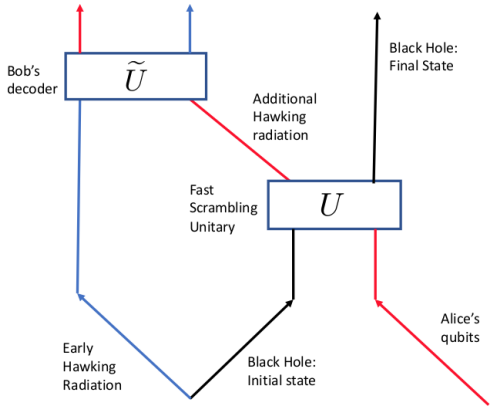


Figure 3.17: Diagrammatic representation of the Hayden-Preskill experiment.

unitary. Then, using the knowledge of the microstate of the black hole, [30] argued that after the half-way point, Bob's decoder will need to collect only an order $\mathcal{O}(1)$ (compared to entropy) number of late Hawking quanta to be able to reconstruct Alice's message. While their argument shows that there exists in principle a quantum operation that can reconstruct the message, Hayden and Preskill did not provide a constructive algorithm to realize the decoding operation.

In [9], it was pointed out that this quantum operation can be realized as follows: we imagine that Bob collects the early radiation emitted by the black hole and he collapses it to form a second black hole. This results in two black holes which are approximately in a maximally entangled state. By applying a complicated unitary on the second black hole, Bob can bring the state of the two black holes in approximately a thermofield-like state. Then the corresponding geometry is that of the two-sided black hole with a non-traversable wormhole. We think of the original black hole as corresponding to the right side while the new black hole to the left. Alice's message can be thought of as throwing particle on the right side of the eternal black hole created by acting on the TFD state with the unitary $e^{i\phi_R(t_0)}$. The step in the Hayden-Preskill protocol of collecting a few more Hawking particles can be thought of as corresponding to the Gao-Jafferis-Wall double trace perturbation, which modifies the state by $e^{ig\mathcal{O}_L\mathcal{O}_R}$. Then Alice's message emerges on the left side in geometric form, where it can be detected by Bob using an operator ϕ_L. This provides us with a specific implementation of the Hayden-Preskill decoding.

3.5.2 Information Recovery Using the Mirror Operators

The question that was addressed in the Hayden-Preskill protocol was how to extract information which fell into the black hole, if we only have access to the Hawking radiation. A somewhat different question is how to extract information which fell into the black hole, if we also have access to the microscopic Hilbert space of the black hole. Of course in that case it is obvious from unitarity that information can in principle be recovered at any time, even before the Page time.

We will present one particular protocol, using the mirror operators, which can be used to recover information from a black hole. If this protocol is applied to a large black hole in AdS, or a black hole in flat-space before Page time, it allows information extraction

provided that one has access to the Hilbert space of the black hole. In AdS this means access to the full Hilbert space of the CFT. On the other hand if the protocol is applied to a black hole in flat space and after Page time then it becomes the analogue of the Hayden-Preskill protocol as formulated in [9] and in particular it allows information recovery purely from the Hawking radiation.

We start with a black hole in AdS dual to a microstate $|\Psi_0\rangle$. At some time $t \approx -t_S$ (here t_S is scrambling time), we throw a qubit into the black hole. This qubit is created in the bulk by acting with the CFT operator $U_\epsilon = e^{i\epsilon\phi(t_0)}$ (appropriately smeared). We wait until the particle has been absorbed, and then we ask what is the CFT operator we need to measure in order to extract the quantum information of the qubit.

Of course in principle the boundary observer, who has access to the microstate of the CFT, can extract the information at any moment in time after the qubit has been injected[25]. Here we present one particular way of extracting the information, which will allow us —in a different limit — to make contact with the Hayden-Preskill protocol.

According to the previous discussions, one natural way to extract the information is the following: after sending in the qubit at $-t_S$, we perturb the CFT Hamiltonian at $t = 0$ by an interaction of the form $U_g = e^{ig\mathcal{O}\widetilde{\mathcal{O}}}$, then two negative shockwaves will be produced, as discussed in section 3.2. The infalling particle collides with one of the shockwaves and undergoes a time-advance pushing it into the "space of the mirrors". It can then be measured by the mirror operator $\widetilde{\phi}(t)$ after scrambling time. All in all, the result of this measurement is captured by the correlator

$$\mathcal{C}_2' \equiv \langle\Psi_0|\, U_\epsilon^\dagger\, U_g^\dagger\, \widetilde{\phi}(t)\, U_g\, U_\epsilon\, |\Psi_0\rangle, \tag{3.136}$$

which according to the discussion of section 3.2 will allow us to measure the qubit at time $t \approx t_S$.

What we have thus is a protocol which allows us to extract the quantum information of a particle which has crossed the horizon, in a time scale of the order of scrambling time. The protocol can be applied to black holes in flat space even before page time, but it requires access to the Hilbert space of the black hole. The reason is that before the Page time the black hole represents the largest part of the Hilbert space, and the mirror operators are supported on it.

Now consider a black hole in flat space, which is after its Page time. The region near the horizon is still approximately thermal, as the time-scale for evaporation is much longer than the timescales (of order scrambling time) relevant for our problem. We can still define the mirror operators for the exterior modes near the horizon. In this limit we have to work with the modular Hamiltonian of the entire system, defined by the Tomita-Takesaki construction, which will not be as simple as the CFT Hamiltonian in the case of a large AdS black hole. After Page time the early radiation represents the largest part of the Hilbert space, hence the mirror operators are mostly supported on the early radiation[26]. Hence, after Page time, implementing the analogue of the state-dependent "double trace" protocol using the mirror operators allows us to extract the information from the cloud of

[25]For example, if the boundary observer applies a time-reversal operator to the state and then evolves forward in time, then the particle will simply pop out of the black hole. Or, relatedly, we can extract the information at a later time by measuring the precursor of the operator which created the particle. We would like to thank J. Maldacena for discussions on this.

[26]While the mirror operators are supported on the early radiation, they can simultaneously play the role of the interior modes behind the horizon, .

Hawking radiation. In particular it is a realization of the Hayden-Preskill protocol in the form discussed in [9].

State-dependence and the Hayden-Preskill protocol

The original Hayden-Preskill protocol to extract quantum information from a black hole, figure 3.17, can only be applied after the Page time. This is because the decoder needs to have access to the larger part of the full Hilbert space, as well as the knowledge of the microstate of the black hole. In particular, in Hayden-Preskill the decoder is state-dependent. When we consider the realization of the protocol in two-sided black hole, as described in [9], the decoder includes the double-trace unitary $e^{ig\mathcal{O}_L\mathcal{O}_R}$, applied *after* we have brought the original state of the system in the the thermofield state, which is done by a state-dependent unitary rotation. Thus this realization of the Hayden-Preskill protocol in [9] is state-dependent as well. In the one-sided realization of the protocol that we proposed, the decoder is a function of the mirror operator $\widetilde{\mathcal{O}}$, which is also state-dependent. This is consistent with the general expectation that the decoder needs to be state-dependent.

Comments on entanglement

The Hayden-Preskill protocol relies on the fact that an old black hole is maximally entangled with the early radiation. This corresponds to an amount of entanglement of $O(S)$ between two tensor factors. In our case we can think of the code-subspace as made out of two tensor factors, corresponding to the algebras $\mathcal{A}$ and $\mathcal{A}'$

$$\mathcal{H}_{|\Psi_0\rangle} = \mathcal{H}_{\mathcal{A}} \otimes \mathcal{H}_{\mathcal{A}'}. \tag{3.137}$$

The state of the black hole $|\Psi_0\rangle$ can be written as an entangled state, similar to the TFD state, in the factors $\mathcal{H}_{\mathcal{A}}, \mathcal{H}_{\mathcal{A}'}$. Our version of the protocol involves information transfer from the tensor factor of $\mathcal{A}$ to $\mathcal{A}'$. A natural question is what is the amount of entanglement between these two algebras. This questions depends on how exactly the algebras $\mathcal{A}$ are defined and how exactly we introduce a cutoff in the number of operators that can be multiplied together in $\mathcal{A}$. We may be able to extend the size of $\mathcal{A}$ to the point where the entanglement entropy between $\mathcal{A}$ and $\mathcal{A}'$ is of order $O(S)$. The algebra $\mathcal{A}$ can be extended as long as the state $|\Psi_0\rangle$ remains a separating vector i.e. it cannot be annihilated by the algebra $\mathcal{A}$. For example in the SYK model we noticed that we can include in $\mathcal{A}$ the fundamental fermions, which would result in an entanglement entropy of $O(S)$.

Irrespective of the fact that we could in principle enlarge $\mathcal{A}$ to attain $O(S)$ entropy, we notice that the decoding protocol may be performed even while keeping the size of these algebras to be $O(1)$. The reason that this is not inconsistent with the $O(S)$ entanglement needed for the Hayden-Preskill protocol is that the code subspace is a very special subspace of the system where the interesting dynamics takes place. In particular it is a state-dependent subspace. For example even in the case of the eternal black hole, while the microscopic entanglement between the two CFTs is $O(S)$, the Gao-Jafferis-Wall protocol can be described within the code subspace corresponding to effective field theory excitations in the bulk which has smaller amount of entanglement.

Finally, in the extrapolation of our experiment to old black holes in flat space, the statement that the mirror operators are supported on the early radiation relies on the fact that the early radiation is maximally entangled with the remaining black hole, with an entanglement of order $O(S)$.

Quantum Teleportation

The Hayden-Preskill experiment, and its two-sided analogues that we have discussed above, are related to quantum teleportation [9]. In the standard teleportation, given two maximally entanglement systems, one wishes to "teleport" a message from one system to another. To do this, one first measures the message (some qubit) and one of the systems such that it is projected into a Bell state. Classically communicating the result of the measurement from system one to two is then used to decode the message. Note that the message need not be physically transported.

The double-trace protocol does not have this two-step measurement process. However, as discussed in [9], the double trace protocol can be related to quantum teleportation. Notice that the the double-trace perturbation can be written as

$$e^{i\,g\mathcal{O}\,\widetilde{\mathcal{O}}} = \frac{1}{2\pi g}\iint d\lambda\, d\widetilde{\lambda}\; e^{ig\lambda\,\mathcal{O}}\; e^{ig\widetilde{\lambda}\,\widetilde{\mathcal{O}}}\, e^{-ig\lambda\widetilde{\lambda}}, \tag{3.138}$$

where we used that the operators $\mathcal{O}$ and $\widetilde{\mathcal{O}}$ commute. Using this identity and equation (3.137), we can calculate the reduced density matrix that corresponds to the mirror operators $\widetilde{\mathcal{O}}$ after the double-trace perturbation. This is easy to do by integrating over the λ variables. One can then show that the result is equal to a reduced density matrix one gets after measuring $\mathcal{O}$ first and then acting with a unitary that depends on the outcome of the measurement on the space of $\widetilde{\mathcal{O}}$'s. This latter operation is like the standard quantum teleportation. Thus the double-trace perturbation in our one-sided setup can be interpreted as a quantum teleportation. This allows us to extract the information $\phi(-t_*)$ by measuring the operator $\widetilde{\phi}(t_*)$.

We now draw a circuit diagram to represent the information transfer during the decoding protocol based on the mirror operators. It is shown in figure 3.18. In this circuit diagram, time flows upwards. The dot on the bottom left denotes the typical pure state $|\Psi_0\rangle$. The Hilbert space of low-energy excitations on top of this state approximately factorizes into ordinary excitations and those behind the horizon of the black hole. As seen in the correlator $\mathcal{C}_2$, we then create a particle ϕ at time $t = -t_*$. This is represented by the input of the state $|\phi\rangle$ in the bottom right part of the figure. This state is part of the Hilbert space spanned by action of algebra $\mathcal{A}$. Let U denote the unitary that acts on this Hilbert space. This unitary rotates the state $|\phi\rangle$ into other states in the same Hilbert space. In physical terms, the state $|\phi\rangle$ undergoes time evolution in the CFT. The goal then is to reconstruct the state $|\phi\rangle$. The action of the double-trace operator unitary V couples simple operators to their mirrors. In the circuit diagram, $\widetilde{U}$ represents the unitary that denotes time evolution in the mirror Hilbert space. This operator acts on the state $|\phi\rangle$ and mixes it with generic mirror excitations like $|\widetilde{\phi}\rangle$. One has access to the mirror excitations (in the code subspace) and in principle one can reconstruct the state $|\phi\rangle$. For this, one first evolves the state $|\phi\rangle$ with the mirror unitary $\widetilde{U}$ for a scrambling amount of time. And then one measures an appropriate mirror particle $|\widetilde{\phi}\rangle$ which the state $|\phi\rangle$ would have the most support on. This is represented by the final steps in the correlator $\mathcal{C}_2$. In the end, we are able to reconstruct the state $|\phi\rangle$.

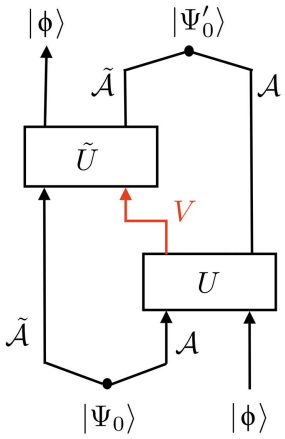


Figure 3.18: An analogue of the Hayden-Preskill protocol: the code subspace approximately factorizes into a tensor product corresponding to the algebras $\mathcal{A}$, $\tilde{\mathcal{A}}$. These tensor factors are entangled and provide the reservoir of EPR pairs needed to perform the teleportation. Here $U, \tilde{U}$ is time evolution in the CFT and $V \approx \mathcal{O}\tilde{\mathcal{O}}$ denotes the perturbation of the CFT Hamiltonian. The figure describes the part of the information flow that is relevant for reconstructing an infalling state $|\phi\rangle$ through measurements behind the horizon. A mirrored diagram exists for the opposite process in which a state is sent from the inside to the outside.

Discussion

In this chapter we discussed the bulk geometry dual to a typical CFT microstate. We proposed that the geometry corresponds to the extended AdS-Schwarzschild solution, including part of the left region. We argued that the existence of state-dependent CFT operators, the mirror operators, supports this idea.

We formulated a one-sided analogue of the Gao-Jafferis-Wall protocol to probe the geometry beyond the exterior, which shows that perturbation of the CFT Hamiltonian by state-dependent operators would allow particles in the region behind the horizon to escape the black hole and get detected in the CFT. This protocol directly relates the smoothness of typical black hole microstates to the technical conjecture that correlators in the thermal ensemble and correlators in the microcananonical ensemble are the same at leading order. The non-trivial aspect of the condition is that it must also hold at timescales of the order of the scrambling time. We looked at various aspects of this conjecture with techniques from statistical mechanics, the Eigenstate Thermalization Hypothesis, chaotic correlators, and numerics in the SYK model to provide partial evidence for the conjecture. We were, however, unable to find a full proof. It would be interesting to find a CFT proof of this conjecture, which would provide additional new evidence in favor of a smooth horizon for typical states.

An important question in understanding the black hole interior in AdS/CFT is to identify a precise CFT computation, whose result contains information about the spacetime behind the horizon. The mirror operators allow us to provide a description of the interior, however the mirror operators are *defined* in such a way that they reproduce a smooth interior. In order to have independent evidence it would be desirable to identify computations

sensitive to the smoothness of the horizon, which involve only ordinary state-independent CFT operators. In this work we have made a small step in this direction, by arguing that the existence of a smooth horizon for a typical state implies certain predictions for OTOCs of ordinary single trace operators on typical states. In particular it implies that they remain close to thermal correlators even at scrambling time, when chaotic effects amplify $1/N$ corrections. As we mentioned, this condition is a *necessary condition* for a smooth horizon and it is formulated as a purely CFT statement which can in principle be investigated without reference to the bulk dual.

This chapter highlights the usefulness of considering state-dependent perturbations of the CFT Hamiltonian, which allows us to probe the region behind the horizon. Similar state-dependent perturbations, but in a class of a-typical states, have been previously considered in [24, 31, 32, 33]. It would be interesting to understand better the possible relevance of such perturbations for general thermal systems, not necessarily dual to black holes in AdS. The perturbations considered in this chapter have the form $\mathcal{O}\widetilde{O}$, connecting the interior to the exterior. It would be interesting to understand possible connections with the proposal formulated in [59].

Finally, it would be interesting to connect the results of this chapterr to the question of how the infalling observer can use the state-dependent operators to experience a smooth interior.

A paper [60] relating the main conjecture of this chapter was published several months after the content of this chapter was published. Of particular interest is the behaviour of pure states in the Schwarzian theory. It was proven that ETH is applicable for these states. Moreover, it was shown that ETH can be extended to more complicated correlators including out-of-time-ordered correlators, thereby proving the conjecture made in this chapter.

There are two assumptions underlying these results. Firstly, it is assumed the Virasoro identity block dominates, which is equivalent to requiring that gravity dominates in the interaction. Secondly, only the diagonal matrix elements were analytically studied. The off-diagonal elements need to be suppressed in order for the conjecture to hold. This was shown analytically in the conformal limit [61] and numerically [45], but not in the general case.

That the Virasoro identity block dominates, or gravity dominates, was already an assumption made in the calculation made in the TFD case and does not place an additional burden on the results of this chapter. In the Schwarzian theory, we can replace the conjecture that correlators are the same to leading order in the canonical and microcanonical ensemble including correlators with time separations of the order of the scrambling times with the conjecture that the off-diagonal terms are suppressed.

Chapter 4

State dependence as a single parameter

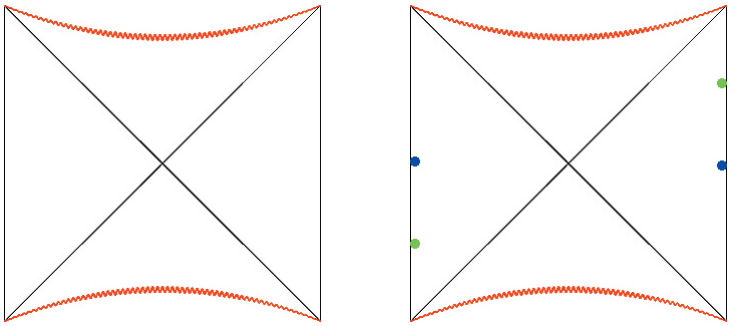
The quantum mechanical behavior of black holes is an ongoing topic started by Hawking [62]. The Firewall paradoxes of AMPS [26, 12, 13] have formalized the problems surrounding the quantum mechanics of black holes. Moreover, these problems persist for large black holes in AdS. The AdS/CFT correspondence [63], therefore, provides a powerful tool to study this topic.

The naive conclusion of these paradoxes is that there can be no operators in the CFT describing the interior of the black hole, and that, therefore, the horizon of the black hole is not smooth. It is, however, possible to construct interior operators that depend on the state of the black hole [6, 7, 8], and thus have a smooth horizon. The geometry described by these operators contains part of the extended AdS-Schwarzschild geometry, including part of the region beyond the interior [2, 3]. In this region it is obvious that time moves opposite to the Killing isometry. The interior operators must, therefore, be explicitly time dependent.

In this chapter, we will examine whether the state-dependence of the interior operators can be captured in a simple form by time dependence. State dependence can directly be rewritten as time evolution in the case of the thermofield double state and its time-shifted cousins [15, 1]. This is possible, more generally, for any ergodic system, because most states will become equal to the other states under time evolution for an ergotic system. State dependence is, therefore, equal to waiting the appropriate amount of time. We explore whether something similar can happen for typical pure state large black holes in AdS, i.e. whether explicit time dependence is enough to avoid the firewall paradoxes.

We also investigate a candidate construction for the interior operators. Tomita-Takesaki theory was used as a motivation for the construction of the state-dependent interior operators. We continue on this path by using the natural cone, described by Tomita-Takesaki theory, which has the elegant property that the interior operators are the same for all states in the natural cone. We investigate whether the natural cone together with explicit time dependence is enough to describe the interior operators of most typical states.

This chapter is organized as follows: In section 4.1, we will discuss explicit time dependence in the case of the thermofield double state, the basics of Tomita-Takesaki theory, and the construction of the state-dependent interior operators called the mirror operators. Next, in section 4.2, we will examine the various firewall paradoxes and see how explicit



(a) (b)

Figure 4.1: a) The Penrose Diagram of the eternal black hole. b) The blue dots correspond to operators which are entangled. An operator at a later time, green dot, is entangled with an earlier operator on the other side.

time dependence can avoid them. Finally, in section 4.3, the natural cone is proposed as a candidate construction.

4.1 Explicit time dependence

Explicit time dependence is needed to get a consistent description of the black hole interior. In the following section, we will show this for the eternal black hole and typical pure state black holes. We will also discuss the Tomita-Takesaki construction used to describe the interior operators.

4.1.1 Eternal Black Hole

The eternal black hole is proposed [17] to be dual to two CFTs in a specific entangled state, the thermofield double state

$$|\Psi_{\text{TFD}}\rangle = \sum_i \frac{e^{-\beta E_i/2}}{\sqrt{Z}} |E_i\rangle_L |E_i\rangle_R , \tag{4.1}$$

where we denote the two CFTs with (L) left and (R) right, and β is the inverse temperature of the black hole. The geometry corresponding to this sate is depicted in figure 4.1a. The geometry has a Killing isometry corresponding to $H_R - H_L$, which flows up on the right part of the geometry and down on the left part of the geometry. This causes the left operator that is entangled with a right operator to move down when we move the right operator up, as depicted in figure 4.1b.

To avoid closed timelike curves, we need to impose that time evolution is generated by $H_R + H_L$, i.e. up in both the left and right side of the geometry. This has consequences for left-right correlators,

$$\langle \mathcal{O}_L(t_1)\mathcal{O}_R(t_2)\rangle = f(t_1 + t_2), \tag{4.2}$$

where f is some function of $t_1 + t_2$. This is different from the one-sided two-point function, which is a function of $t_1 - t_2$. Left and right operators commute, even though the left-right two-point function is non-zero. This is because the two CFTs are causally disconnected. However, by considering a double trace perturbation [5, 9], of the form $\delta H = g\mathcal{O}_L\mathcal{O}_R$, a

message can be send from one CFT to the other. This provides evidence for the smoothness of the black hole horizon in this state, as such a probe crosses the horizon.

There is a class of states [15] that have the same entanglement structure called the time-shifted thermofield double states,

$$|\Psi_T\rangle = \sum_i \frac{e^{-\beta E_i/2}}{\sqrt{Z}} e^{iE_i T} |E_i\rangle_L |E_i\rangle_R . \tag{4.3}$$

These states can also be made traversable [1], i.e. send a message from one side to other. However, the protocol depends on T and is, therefore, state dependent with a single parameter.

4.1.2 Tomita-Takesaki Theory

The construction developed by Tomita and Takesaki state that for a given algebra $\mathcal{A}$ and reference state $|\psi\rangle$, which obeys

- $|\psi\rangle$ is cyclic, i.e. $\mathcal{A}|\psi\rangle$ spans the entire Hilbert space,
- $|\psi\rangle$ is seperating, i.e. $A|\psi\rangle = 0$ only for $A = 0$,

then the commutant $\mathcal{A}'$ of the algebra $\mathcal{A}$ can constructed. This is done by using the Tomita operator S as follows

$$\begin{aligned} SA|\Psi_0\rangle &= A^\dagger|\Psi_0\rangle, \\ \Delta = S^\dagger S\,, \qquad S &= J\Delta^{1/2}\,, \\ \widetilde{\mathcal{O}} = J\mathcal{O}J\,, \quad \mathcal{A}' &= J\mathcal{A}J\,, \end{aligned} \tag{4.4}$$

where J is an anti-unitary operator called *modular conjugation*, while Δ is called the *modular operator*. An in-depth discussion about the properties of Tomita-Takesaki theory can be found in [64].

It is useful to put this construction in the context of the thermofield double, where we have the following identities

$$\begin{aligned} \mathcal{O}_L &= J\mathcal{O}_R J, \\ \Delta &= e^{-\beta(H_R - H_L)}. \end{aligned} \tag{4.5}$$

Here, we notice that the modular operator Δ is a function $H_R - H_L$, not $H_R + H_L$. This means that we can generate time evolution on the right exterior with the modular operator by using Δ^{is}, but we need to use explicit time dependence, to compensate for the minus sign, if we also use the modular operator on the left exterior in order to keep a consistent causal structure. This may seem unnecessarily complicated for the thermofield double, however, this is unavoidable in the case of a typical state.

The algebra and its commutant remain the same in the case of the time-shifted thermofield double states. This, however, means that left-right correlation functions change as function of the timeshift. It would, therefore, be necessary to work with precursors that absorb this timeshift to get the same correlation functions as in the situation without timeshift. For example, if we want the same correlators between left and right operators in the thermofield double state and a time-shifted thermofield double state

$$\langle\Psi_{\text{TFD}}|\,\mathcal{O}_L(t_1)\mathcal{O}_R(t_2)\,|\Psi_{\text{TFD}}\rangle = \langle\Psi_T|\,X_L(t_1)\mathcal{O}_R(t_2)\,|\Psi_T\rangle\,, \tag{4.6}$$

we would need to use a precursor X_L on the left side, and identify $X_L(t) = \mathcal{O}_L(t - T)$.

4.1.3 Typical Black Hole

The typical black hole microstate is defined as a superposition of energy eigenstates

$$|\Psi\rangle = \sum c_i \, |E_i\rangle \,, \tag{4.7}$$

where we sum over the energy eigenstates in the window $E_0 \pm \delta E$, and c_i are random complex numbers chosen with the Haar measure. E_0 is given by the mass of the black hole and δE is an order one number making the window wide enough to account for the entropy of the black hole. Black holes of this type are not formed by normal collapse [25].

We want to use the Tomita-Takesaki construction to describe the interior of these black holes. However, typical states are not cyclic and seperating. Nonetheless, they are almost cyclic and almost seperating, i.e. it is difficult to annihilate these states and perturbations around the black hole can be well described by low-point correlation functions. We can, therefore, use the mirror operators constructed in [6, 7, 8].

We can use a truncated algebra $\mathcal{A}$ of single-trace operators in frequency space, defined by

$$\mathcal{A} = \text{span}\left\{\mathcal{O}_{\omega_1}, \mathcal{O}_{\omega_1}\mathcal{O}_{\omega_2}, ..., \mathcal{O}_{\omega_1}....\mathcal{O}_{\omega_n}\right\}, \tag{4.8}$$

where $n \ll N$. This linear set is only approximately an algebra as we demand $n \ll N$. Moreover, we will have to work with coarse grained operators to avoid problems caused by the discrete nature of the energy levels.

$$\begin{aligned} \mathcal{O}_\omega^{\text{exact}} &\equiv \int_{-\infty}^{+\infty} dt \; e^{i\omega t}\mathcal{O}(t). \\ \mathcal{O}_\omega &\equiv \frac{1}{\sqrt{\delta\omega}} \int_{\omega}^{\omega+\delta\omega} \mathcal{O}_{\omega'}^{\text{exact}} d\omega', \end{aligned} \tag{4.9}$$

where now the set of allowed ω's is discretized with step $\delta\omega$. In (4.9) we have divided by $\sqrt{\delta\omega}$ in order to have an operator whose correlators are stable under small changes of the bin size $\delta\omega$. We also need to impose an upper cutoff in the allowed frequencies $|\omega| \leq \omega_*$. The reason is that the mirror operators are meaningful when the small algebra cannot annihilate the state. In a thermal state we find that $\langle \mathcal{O}_\omega^\dagger \mathcal{O}_\omega \rangle \propto e^{-\beta\omega}$. For large ω this is extremely close to zero, implying that the operator $\mathcal{O}_\omega$ almost annihilates the state.

The limitations for the algebra are chosen in such a way that typical states are not annihilated by the algebra, and typical states are, therefore, seperating.

The algebra forms, when acting on the reference state, the small Hilbert space

$$\mathcal{H}_{|\Psi_0\rangle} = \mathcal{A}\,|\Psi_0\rangle \,, \tag{4.10}$$

in which physics is described with the black hole as a background. It does describe objects falling into the black hole, but not large perturbations to the black hole, such as black hole mergers for example. The typical state is cyclic by construction with respect to the small Hilbert space. Therefore, we can make use of Tomita-Takesaki theory. Moreover, using large N factorization and the KMS condition relevant for equilibrium states, it is possible to show [8] that at large N the CFT Hamiltonian acts on the code subspace similar to the (full) modular Hamiltonian

$$\Delta = \exp[-\beta(H - E_0)] + O(1/N). \tag{4.11}$$

From (4.4),(4.11) follows that at large N the mirror operators are defined by the equations

$$\begin{aligned}\widetilde{\mathcal{O}}_\omega|\Psi_0\rangle &= e^{-\frac{\beta H}{2}}\mathcal{O}_\omega^\dagger e^{\frac{\beta H}{2}}|\Psi_0\rangle,\\ \widetilde{\mathcal{O}}_\omega\mathcal{O}_{\omega_1}...\mathcal{O}_{\omega_n}|\Psi_0\rangle &= \mathcal{O}_{\omega_1}...\mathcal{O}_{\omega_n}\widetilde{\mathcal{O}}_\omega|\Psi_0\rangle,\\ [H,\widetilde{\mathcal{O}}_\omega]\mathcal{O}_{\omega_1}...\mathcal{O}_{\omega_n}|\Psi_0\rangle &= \omega\widetilde{\mathcal{O}}_\omega\mathcal{O}_{\omega_1}...\mathcal{O}_{\omega_n}|\Psi_0\rangle.\end{aligned} \tag{4.12}$$

The last line generalizes to higher powers of the Hamiltonian H, even though H is not in the small algebra.

Specifying the action of the operators on the small subspace $\mathcal{H}_{|\Psi_0\rangle}$, as in (4.12), is not sufficient to know the time when the operators act. We also need to specify a *time ordering* between them and with the normal operators. To describe effective field theory in the bulk, we must define the mirror operators to go against the killing isometry in the bulk. This does not fully fix the time evolution as an overall shift remains. This shift is similar to the relation between the thermofield double and its time-shifted cousins. The one-parameter family, labeled by T, of possible choices for how to localize the mirrors in physical time is given by

$$\widetilde{\mathcal{O}}_T(t) = \int_{-\omega_*}^{\omega_*} d\omega\ e^{-i\omega(t-T)}\widetilde{\mathcal{O}}_\omega, \tag{4.13}$$

where t labels the physical CFT time at which the operator is localized. Using this Fourier transform, we can setup the mirror operator equations in position space.

$$\begin{aligned}\widetilde{\mathcal{O}}_T(t)\,|\Psi_0\rangle &= e^{-\frac{\beta H}{2}}\mathcal{O}^\dagger(T-t)e^{\frac{\beta H}{2}}\,|\Psi_0\rangle\,,\\ \widetilde{\mathcal{O}}_T(t)A(t_1,t_2,...)\,|\Psi_0\rangle &= A(t_1,t_2,...)\widetilde{\mathcal{O}}_T(t)\,|\Psi_0\rangle\,,\\ [H,\widetilde{\mathcal{O}}_T(t)]A(t_1,t_2,...)\,|\Psi_0\rangle &= A(t_1,t_2,..)e^{-\frac{\beta H}{2}}[H,\mathcal{O}^\dagger(T-t)]e^{\frac{\beta H}{2}}\,|\Psi_0\rangle\,.\end{aligned} \tag{4.14}$$

These statements are only approximately true and not completely local, as we have binned the Fourier modes and imposed cut-off frequencies. The free parameter T is needed to get a consistent picture for when the mirror operators are defined at different times, i.e. one person uses $|\Psi_0\rangle$ the state, but someone else might set up the experiment some time later and use $e^{iHt}\,|\Psi_0\rangle$. They must choose T carefully to get consistent results, as the mirror operators they naively obtain – without T – are not the same but related as precursors to each other. Normally, it is the complexity that tells you whether you are using the proper operator[1] or a precursor. We will suppress writing the subscript T for the case $T=0$, such that $\widetilde{\mathcal{O}}_0 = \widetilde{\mathcal{O}}$, for notational ease. The results for the case of general T can be obtained as precursors of the $T=0$ case.

From the first line of equation (4.14), it is clear that the mirror operators move backwards in time under normal Hamiltonian evolution. This can also be seen from third line of equation (4.12), where the sign is different from what one expects for a normal operator. The mirror operators must, therefore, be explicit time dependent to ensure that they are forward moving in time and to obtain a consistent causal structure.

The geometry of typical state black holes is conjectured [2, 3] to include part of the left exterior of the extended Penrose diagram, as depicted in figure 4.2. This part of the geometry is described by the mirror operators.

[1] For a bulk picture interpretation one wants to use the simplest operator. Mathematically speaking all precursors are proper operators.

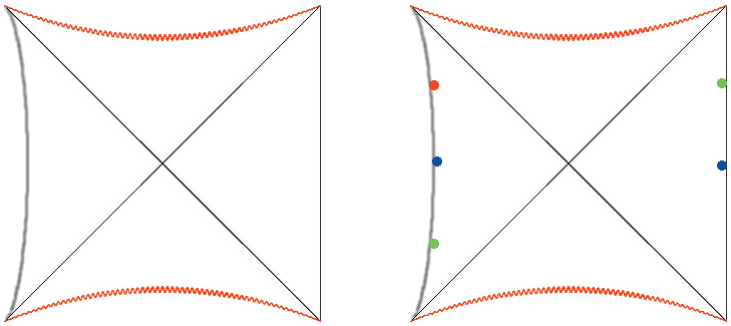


Figure 4.2: a) Penrose diagram of the typical black hole. b) Blue dots denote entangled operators, and green dots denote entangled operators at a later time, similar to the eternal black hole. However, someone using the state at the later time as the reference state for the mirror operator construction would place the location of the operator that is entangled with the right green operator at the red dot, thereby making the construction time-dependent.

4.2 Avoiding the paradoxes

We have seen that explicit time dependence is a consequence of physical requirements, i.e. to have consistent bulk picture. Explicit time dependence encodes some information about the state in the operators used to construct the interior of a black hole. They are, therefore, state-dependent in this simple form. This leads us to the following claim:

Explicit time dependence avoids the firewall paradoxes.

To test the claim, we only need to study a set of states C,

$$C = \{|\Psi_1\rangle, |\Psi_2\rangle, \dots\}, \tag{4.15}$$

where the states $|\Psi_i\rangle, |\Psi_j\rangle$ do not evolve into each other under time evolution, and consider the interior operators as fixed (state-independent) for this set of states. Each state drawn from the Haar measure, for a given energy window, should be part of this set at some time. The interior operators for a given state $|\Psi(0)\rangle$ are thus obtained by finding the time when $|\Psi(t)\rangle$ is part of the set C and use explicit time evolution with the fixed interior operators to obtain the interior operators for $|\Psi(0)\rangle$.

In this section we discuss the various paradoxes [12, 8], and how they are resolved by explicit time dependence.

4.2.1 The "$N_a \neq 0$" paradox

These arguments [13, 27] support the idea that the horizon is not smooth by comparing the excitations in the modes of the Hawking radiation with the modes of an infalling observer.

4.2.1.1 The Paradox

The number operator, $N_b = b^\dagger b$, measures the number of particles at frequency ω seen by the outside observer. A different number operator,

$$N_a = \frac{1}{1-e^{-\omega\beta}}\left[\left(b^\dagger - e^{-\omega\beta/2}\widetilde{b}\right)\left(b - e^{-\omega\beta/2}\widetilde{b}^\dagger\right) + \left(\widetilde{b}^\dagger - e^{-\omega\beta/2}b^\dagger\right)\left(\widetilde{b} - e^{-\omega\beta/2}b^\dagger\right)\right], \tag{4.16}$$

measures the number of particles seen by the infalling observer. It should be obvious that, to leading order, the following holds

$$[H_{\text{CFT}}, N_b] = 0, \tag{4.17}$$

which means that the Hamiltonian and this number operator can be simultaneously diagonalized, at least to leading order.

$$\begin{aligned} H_{\text{CFT}} \left|E_i, n_{i,b}\right\rangle &= E_i \left|E_i, n_{i,b}\right\rangle, \\ N_b \left|E_i, n_{i,b}\right\rangle &= n_{i,b} \left|E_i, n_{i,b}\right\rangle. \end{aligned} \tag{4.18}$$

This leads to a contradiction with the idea that the horizon is smooth when we assume that N_a is a fixed state-independent operator.

$$\langle N_a \rangle = \text{ Tr}[\rho_m N_a]/\mathcal{N} = \left\langle E_i, n_{i,b}\right| N_a \left|E_i, n_{i,b}\right\rangle = O(1), \tag{4.19}$$

where ρ_m is microcanonical density matrix, and $\mathcal{N} = \text{ Tr}[\rho_m]$ is the normalization. The last equality holds, because N_a is a positive operator and any state with $\langle N_a \rangle = 0$ has a thermal expectation value of N_b and, therefore, the expectation value of N_a must be order one in eigenstates of N_b.

4.2.1.2 The Resolution

It is important that correlators of typical states are close to correlators in the microcanonical ensemble for the previous argument. This is obviously not true in the case of state-dependent operators. In our case the operators are not fully dependent on the state, only on the time. Therefore, it not obvious that the arguments of the paradoxes fail for explicit time dependence.

For correlators with normal operators averaging over the typical states should give the correlator in the microcanonical ensemble [49],

$$\int [d\mu] \left\langle \Psi\right| A \left|\Psi\right\rangle = \text{ Tr}[\rho_m A]/\mathcal{N}, \tag{4.20}$$

Where A is some hermitian operator. We can estimate that each individual typical state is close to microcanonical ensemble by calculating the variance. These calculations are done by averaging over unitaries that rotate the state. This causes the crossterms with random phases to drop out and the diagonal terms remain, which give the trace.

The measure $[d\mu]$ is used to average over all typical states. It can be decomposed into a part that averages over the size of coefficients c_i of the state, and a part that averages over the phases of the state, $[d\mu] = [d\hat{\mu}]\frac{dt}{t_f}$. The second part, averaging over the phases, is equal to averaging over the time evolution of a state. Therefore, we obtain

$$\begin{aligned} \text{Tr}[\rho_m A]/\mathcal{N} &= \int [d\mu] \left\langle \Psi\right| A \left|\Psi\right\rangle, \\ &= \int [d\hat{\mu}] \int_0^{t_f} \frac{dt}{t_f} \left\langle \Psi(t)\right| A \left|\Psi(t)\right\rangle, \\ &= \int [d\hat{\mu}] \langle A \rangle_\Psi, \end{aligned} \tag{4.21}$$

where t_f is the time of a full orbit, and $\langle A \rangle_\Psi$ is the long time average.

We can use the same steps to get an estimate for the variance to conclude that most states in the energy window are close to the microcanonical ensemble. However, when A is partially constructed from interior operators, this construction fails, because the interior operators introduce an explicit time dependence and we cannot use the third line.

The correlator of a time-independent operator is close to the long time average, averaging only over the time of the state. The explicit time dependence of the interior operators means that we cannot use equation (4.19) for a typical state as we cannot do the change of basis while we have both normal and interior operators in the correlator. Correlators with both normal and interior operators are not close to the micro-canonical ensemble, and the paradox is, therefore, avoided.

It is important to note that the interior Fourier modes are sensitive to the explicit time dependence, even though they act like Fourier modes. Firstly, the interior modes only need to act like Fourier modes in the given background of states, i.e. the set of black holes of given mass, not all states in the Hilbert space. Secondly, we must consider the Fourier modes to be smeared to avoid becoming sensitive to the energy levels, which also introduces a time dependence.

4.2.2 Lack of Left Inverse

We now move to another paradox [12].

4.2.2.1 The Paradox

On one hand we know that the interior operator $\widetilde{b}^\dagger_\omega$ acts as a creation operator as we write, to leading order, the following

$$\begin{aligned}\langle\Psi_0|\,\widetilde{b}\widetilde{b}^\dagger - \widetilde{b}^\dagger\widetilde{b}\,|\Psi_0\rangle &= \langle\Psi_0|\,\widetilde{b}e^{\frac{\omega\beta}{2}}b - \widetilde{b}^\dagger e^{-\frac{\omega\beta}{2}}b^\dagger\,|\Psi_0\rangle\,,\\ &= \langle\Psi_0|\,bb^\dagger - b^\dagger b\,|\Psi_0\rangle\,,\\ &= 1,\end{aligned} \tag{4.22}$$

where we used that the interior and exterior operators commute and, secondly, we used $\widetilde{b}\,|\Psi_0\rangle = e^{-\beta H/2}b^\dagger e^{\beta H/2}\,|\Psi_0\rangle = e^{-\frac{\omega\beta}{2}}b^\dagger\,|\Psi_0\rangle$. We can, therefore, use $[\widetilde{b},\widetilde{b}^\dagger] = 1$ inside correlators. Moreover, this should hold as an operator statement if the interior operators are state-independent, which means that $\widetilde{b}^\dagger$ must have a left-inverse.

On the other hand, we have that $[H,\widetilde{b}^\dagger] = -\omega\widetilde{b}^\dagger$ and, therefore, it lowers the energy of the state. There are fewer states at lower energies and $\widetilde{b}^\dagger$ must annihilate part of the typical state and, therefore, cannot have a left-inverse. We will see that this does not interfere with C, i.e. the kernel does not need to significantly overlap with C.

4.2.2.2 The Resolution

We can avoid annihilating part of the state by fine-tuning the operators to the phases of the state, or by fine-tuning the phases of the state to the operators. This is equivalent to selecting a fine-tuned time of the state. We can check in a simple toy model whether mapping to a smaller set of states can leave the norm of vector intact. It is good to remember that C is really small compared to the size of the Hilbert space. The phases are fixed this halves the dimensionallity of the space, and flipping the sign of an energy

eigenstate is the same as time evolution. Only one combination of signs is in C, which is a 2^{-e^S} part of the real part of the Hilbert space.

We can use matrices as a simple toy model to test whether we can avoid annihilating part of the state by tuning the phase. For example, consider the following matrix equation: $M\vec{y} = \vec{x}$ where M is an $m \times n$ matrix, $\vec{y}$ is a length n vector, and x is a length m vector. We consider the case where $m > n$ and compare the between y with random complex elements and $\vec{y}$ with positive real elements. To be explicit, we pick $m = 2$ and $n = 1$ and consider $\vec{y}^\dagger M^\dagger M\vec{y}$, thus an example map that we are interested in is $M = \begin{bmatrix}1 & 0\end{bmatrix}$, which gives

$$M^\dagger M = \begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix}. \tag{4.23}$$

It is easy to see that this will partially annihilate the state and reduce the size of the vector to $|\vec{\boldsymbol{x}}|^2 = \frac{1}{2}$, in both the case of the complex vector and of the 'positive' vector. We can, however, change the basis and work with the similar matrix

$$M^\dagger M = \frac{1}{2}\begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix}. \tag{4.24}$$

Now we see that $|\vec{\boldsymbol{x}}|^2 = (y_1^\dagger y_1 + y_2^\dagger y_1 + y_1^\dagger y_2 + y_2^\dagger y_2)/2$. For the complex factors the random phases of the cross terms average out and we are again left with $|\vec{\boldsymbol{x}}|^2 = \frac{1}{2}$. This is different for the 'positive' vector, because there are no phases and nothing cancels. This kind of map can be generalized to higher dimensions, where the elements of $\vec{y}$ become uncorrelated and the norm of the approaches $|\vec{\boldsymbol{x}}|^2 = 1 - \epsilon$, where ϵ is suppressed by the size of the vectors. In higher dimensions this generalizes to

$$\begin{bmatrix}1 & 0 & 0 & \dots & 0\\ 0 & 0 & 0 & \dots & 0\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & 0 & \dots & 0\end{bmatrix} \to \frac{1}{m}\begin{bmatrix}1 & 1 & 1 & \dots & 1\\ 1 & 1 & 1 & \dots & 1\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 1 & 1 & 1 & \dots & 1\end{bmatrix}. \tag{4.25}$$

More generally, most entries of the matrix need to be positive, and that some of its eigenvalues are zero. It is clear that these maps reduce the size of the state, but they do not reduce the norm of the state (in the large N limit).

We can use the same steps to obtain a matrix that reduces the norm of state. For example, we look at the following matrix

$$b = \begin{bmatrix}\sqrt{3}-1 & \sqrt{3}+1 & 2 & 0\\ 1 & -1 & 1 & 3\\ \sqrt{3}+1 & \sqrt{3}-1 & 2 & 0\end{bmatrix}, \tag{4.26}$$

such that

$$\frac{1}{12}b^\dagger b = \begin{bmatrix}1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\end{bmatrix} + \frac{1}{4}\begin{bmatrix}-1 & 1 & -1 & 1\\ 1 & -1 & 1 & -1\\ -1 & 1 & -1 & 1\\ 1 & -1 & 1 & -1\end{bmatrix}. \tag{4.27}$$

The alternating signs reduce the error we make when we use $\frac{1}{12}b^\dagger$ as an approximate left inverse provided that the elements of the vector we act on are positive. We can estimate

	$\lvert 1 - v^\dagger(\frac{1}{12}b^\dagger b)v\rvert$
Free phases	1.0783
Fixed phases	0.0538

Table 4.1: We averaged over 10^6 random draws of the vector v. The phases were fixed by taking the element wise absolute value of the random vector.

the error by drawing random vectors v with the Haar measure and calculate the error we make when we use $\frac{1}{12}b^\dagger$ as a left inverse.

For example, see table 4.1, where we see that we cannot change the basis of the state without also changing the operators, while keeping the result the same, because changing the state would introduce phases. Thus we cannot change the basis of the state while working with interior operators, just as in section 4.2.1.2.

4.2.3 Other Paradoxes

There are several other paradoxes, which we will only discuss briefly as the approaches from the previous resolutions also applies to these.

1. The Strong Subadditivity Puzzle.

For a black hole to have a smooth horizon it is necessary that particles just outside the black hole are entangled with the particles just inside the horizon. Moreover, particles just outside the horizon must be entangled with early radiation for information to escape the black hole [12]. However, monogamy of entanglement forbids this. This assumes that the interior is independent of the exterior, which is something that we did not require, i.e. the interior can be constructed from fine-tuned multi-trace operators, which makes these operators dependent on the exterior.

2. The $[\text{Exterior}, \text{Interior}] \neq 0$ Paradox.

Commutators between an operator and a scrambled version thereof tend to be order one [12].

$$[U\mathcal{O}U^\dagger, \mathcal{O}] \sim O(1), \tag{4.28}$$

for a scrambling unitary U, which should be detectable outside the black hole. However, we do not require that the interior operators are of the form $U\mathcal{O}U^\dagger$ as we allow them to be fine-tuned, which gives the freedom for the interior and exterior operators to commute.

4.3 The natural cone

There is a lot of freedom in choosing the set C, as we can replace each state with the time evolution of that state. The construction of Tomita and Takesaki gives a candidate for the set C, which does not have this ambiguity. The set of states of the form

$$\mathcal{P}_{|\psi\rangle} = \overline{\{AJAJ\,|\psi\rangle : A \in \mathcal{A}\}}, \tag{4.29}$$

is called the natural cone [64] and the states in this cone have the property that modular conjugation J is the same for all cyclic states in the cone. The mirror operators constructed in section 4.1.3 are, therefore, the same for all states inside the natural cone. We will look at some other properties of the natural cone in the next subsection.

We propose to use $C \approx \mathcal{P}_{|\psi\rangle} \cap \{\text{Typical States}\}$, where all states come from the same energy window. The following algorithm can be used to obtain the mirror operators for most typical states from the same energy window.

1. Pick a random typical state $|\psi_0\rangle$ to serve as a reference state.
2. Construct the mirror operators, without timeshift (see section 4.1.3), for $|\psi_0\rangle$.
3. Construct the natural cone $\mathcal{P}_{|\psi_0\rangle}$.
4. Find the time when a test state $|\psi_1(t)\rangle$ is close to the natural cone $\mathcal{P}_{|\psi_0\rangle}$, up to some $O(1/N)$ tolerance.
5. Use the same mirror operators for $|\psi_1(t)\rangle$ and $|\psi_0\rangle$.
6. The mirror operators for $|\psi_1(0)\rangle$ are obtained by evolving from $|\psi_1(t)\rangle$ and taking explicit time dependence into account.
7. Precursors must be used to obtain the correct correlators, similar to equation (4.6).

It is enough for the test state to come close to the natural cone in step 4, because a $O(1/N)$ perturbation of a state will not affect leading order results. Whether a state gets close to the natural cone is discussed in section 4.3.1.2. In section 4.3.2.2, we will discuss that this algorithm is independent of the choice of reference state at leading order.

4.3.1 Basic Properties

The modular conjugation J and modular hamiltonian Δ, also called the modular objects, are defined for a specific state, and may differ for different states. Modular theory, however, tells us that there are some relations between the modular objects of different states [64]. For example, if two states are related by a unitary $|\phi\rangle = U|\psi\rangle$ then the modular object for $|\psi\rangle$ and the algebra $U\mathcal{A}U^\dagger$ is $(U\Delta U^\dagger, UJU^\dagger)$.

Another relation between modular objects can be found when we look at the following set of states.

$$\mathcal{P} = \overline{\{AJAJ|\psi\rangle : A \in \mathcal{A}\}}, \tag{4.30}$$

where $\mathcal{P}$ is called the natural positive cone associated with the pair $(\mathcal{A}, |\psi\rangle)$. The overline denotes the closure in the Hilbert space $\mathcal{H}$. Note that the states in this cone may not be normalized, but normalization does not remove any of its interesting properties. Some of the properties of $\mathcal{P}$ are

- $\mathcal{P} = \overline{\{\Delta^{-1/4}AA^\dagger|\psi\rangle : A \in \mathcal{A}\}}$.
- $\mathcal{P}$ forms a convex cone.
- $\mathcal{P}$ is self-dual.
- $\text{span}(\mathcal{P}) = \text{span}(\mathcal{H})$.
- $AJAJ\mathcal{P} \subset \mathcal{P}, \forall A \in \mathcal{A}$.
- $J|\phi\rangle = |\phi\rangle, \forall |\phi\rangle \in \mathcal{P}$.
- $|\phi\rangle$ is cyclic $\Leftrightarrow$ $|\phi\rangle$ is separating, $\forall |\phi\rangle \in \mathcal{P}$.

Moreover, if $|\phi\rangle \in \mathcal{P}$ is cyclic (and, therefore, separating) then the modular conjugation is the same $J_{|\phi\rangle} = J$ and the natural positive cone is the same $\mathcal{P}_{|\phi\rangle} = \mathcal{P}$.

The fact that $\mathcal{P}$ is self-dual means that the dual of $\mathcal{P}$, defined by

$$\mathcal{P}^D = \{|\psi\rangle \in \mathcal{H} : \langle\psi|\phi\rangle \geq 0, \forall\, |\phi\rangle \in \mathcal{P}\}, \tag{4.31}$$

is equal to $\mathcal{P}$.

We proposed to use $C \approx \mathcal{P} \cap \{\text{Typical States}\}$, where all states come from the same energy window. The main advantage is that the modular operator J is the same in the natural cone and the mirror operators are, therefore, state-independent within the natural cone. It is, thus, important that every typical state gets close to some element from the natural at some point in time,

$$|\Psi_1(t_0)\rangle + |\delta\rangle = a_0 |\Psi_0\rangle + a_1 A_1 \widetilde{A}_1 |\Psi_0\rangle + a_2 A_2 \widetilde{A}_2 |\Psi_0\rangle \ldots \tag{4.32}$$

where $a_i > 0$, and the norm of $|\delta\rangle$ is small compared to the norm of $|\Psi_1(t_0)\rangle$. The operators A_i that are most useful are of a specific form, see equation (4.34), which we will discuss later. The motivation that this superposition results in a small error term comes from the property that the overlap between different typical states is significant [20]. For example

$$\max_t(\langle\Psi_1(t)|\Psi_0\rangle) = \frac{\pi}{4}, \tag{4.33}$$

which will only increase when we add more degrees of freedom, see appendix C.2 for more examples.

4.3.1.1 Restricting to the Small Algebra

Most properties are conserved when we restrict to the energy window. However, it is only self-dual if we discount the part of the dual cone outside the energy window.

The second restriction we must make is restricting to the small algebra. The main property we must test is whether the restricted cone spans the energy window. To do that we estimate how much volume the new cone fills. The easiest way to do this is to count the number of degrees of freedom we have for the construction of the natural cone, while we restrict ourselves with operators $AJAJ$ that stay within the energy window.

A single Fourier mode works as $[H, \mathcal{O}_\omega \widetilde{\mathcal{O}}_\omega] = 0$. This, however, is badly behaved as a state, $\mathcal{O}_\omega \widetilde{\mathcal{O}}_\omega |\Psi\rangle$, for example, is not normalizable. We must smear the Fourier mode to remedy this problem. The width of the frequencies that we smear over, ω_s, must be very small as the smearing causes $\mathcal{O}_\omega \widetilde{\mathcal{O}}_\omega |\Psi\rangle$ to be slightly wider than the energy window. We can view this spillover, or the removal of the spillover, as a perturbation as long as it is small. On the other hand, the smearing cannot be too small, because it should be at least a few level spacings in size, thus $O(e^{-S}) < \omega_s < O(1/N)$.

Moreover, the frequency of the Fourier mode used cannot be too large. Less than some cut-off frequency ω_*, as discussed in 4.1.3. There is some discussion on how large this frequency can be, but we will be conservative in the following estimates.

We can multiply Fourier modes to obtain another operator that fits our requirements. So we can use the following operator

$$A = \mathcal{O}_{\omega_1}^{n_1} \mathcal{O}_{\omega_2}^{n_2} \mathcal{O}_{\omega_3}^{n_3} \ldots \tag{4.34}$$

where we multiplied the operator with frequency ω_i times itself n_i times. All N_i must be smaller than some $n_m = O(N^0)$, as the states we are looking for are almost in thermal

equilibrium and we should not deviate too far from that. We assume that the different frequencies are distinct as we binned them with ω_s, which means that the ordering does not matter. In appendix C.1, we show that states perturbed by single Fourier modes in the discussed manner are independent from each other, even when the frequencies are almost the same.

The number of states obtained is given by the number of different operators we can construct, as detailed above.

$$\#\text{States} = (n_m + 1)^{(n_p \omega_* / \omega_s)}, \tag{4.35}$$

where $n_p = O(N^0)$ is the number light primaries in the theory. The plus one comes from the possibility that a Fourier mode is absent in the operator. This estimate is exponentially large and certainly comparable with 2^n , but it is an overcounting as we have not restricted sum of frequencies to be below the frequency cut-off. This, however, does not change the order of magnitude, as the size of the smearing is a much larger contribution to the number of operators that we can use to construct new states.

These counting arguments are reminiscent of the counting of 't Hooft brick wall model [65], where the correct entropy can be derived from assuming a cut-off provided by a brick wall at the horizon.

We will assume that we can use the properties of the previous subsection, even when using the small algebra. This is, however, the main weakness of this proposal, as there is direct competition between the number of states that we can create in this manner, which we want to be large, and the size of the small Hilbert space, which we want to be small.

4.3.1.2 Going from any State to $\mathcal{P}$

The natural cone $\mathcal{P}$ is a self-dual cone. This means that the inner product between two states, $|\Psi\rangle, |\Phi\rangle$ from the natural cone is positive $\langle\Phi|\Psi\rangle > 0$. It is, however, useful to consider the real subspace $\mathcal{H}_\mathcal{P}$ that contains the natural cone $\mathcal{P}$. This means that the inner product between any state $|\chi\rangle$ from the real subspace $\mathcal{H}_\mathcal{P}$ and any state from the natural cone $|\Psi\rangle$ is real, $\langle\chi|\Psi\rangle \in \mathbb{R}$. We can do this because both $\mathcal{P}$ is convex, and spans the small Hilbert space. We can, therefore, write any vector in $|\chi\rangle \in \mathcal{H}_\mathcal{P}$ as a sum over vectors $\sum_i c_i |\psi_i\rangle$, where $|\psi_i\rangle \in \mathcal{P}$ and $c_i \in \mathbb{R}$. It is useful to consider these real subspaces as it is much more intuitive what the cones look like in a real subspace. We show a low dimensional toy example in figure 4.3.

If we make some simplifying assumptions, we can make an estimate whether a typical state $|\Psi\rangle$ passes through $\mathcal{P}$ under time evolution.

1. $|\Psi\rangle$ passes through $\mathcal{H}_\mathcal{P}$ at independent uniformly random points.

We allow an exponentially small tolerance when we consider whether a state passes through $\mathcal{H}_\mathcal{P}$ at some time. This is necessary because there are, generally speaking, energy levels that are not related by rational numbers. Therefore, it would take an infinite amount of time get a specific phase alignment between the energy levels.

The assumption that the points are uniformly distributed is too strong, as correlations that do exist tend to push the points at which the state passes through $\mathcal{H}_\mathcal{P}$ away from each other. This happens because the inner product $\langle\Psi| e^{-iHt} |\Psi\rangle$ goes to zero very fast, meaning that at large time separation the states are almost orthogonal.

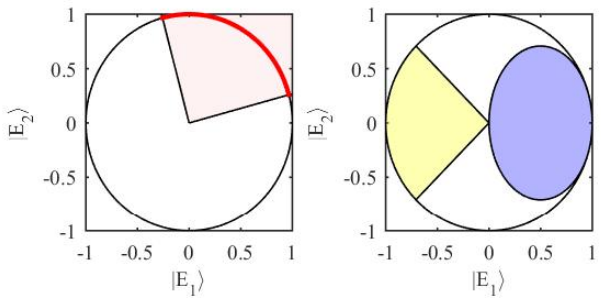


Figure 4.3: Left) We start with $\mathbb{C}^2$ and choose two random orthogonal vectors in this space. This spans a real subspace $\mathbb{R}^2$, and positive superpositions of the two vectors form a self-dual cone. The red circle arc is the unit normalized part of the self-dual cone. Right) The three dimensional analogue, where we show the upper hemisphere and the self-dual orthant (yellow) and circular (blue) cones.

This means that at each pass through $\mathcal{H}_\mathcal{P}$ the state has a probability P_1 to be in $\mathcal{P}$, which is given by the ratio of the volumes

$$P_1 = \frac{\text{Vol}(\mathcal{P} \cap E_\delta \cap S^{n-1})}{\text{Vol}(\mathcal{H}_\mathcal{P} \cap E_\delta \cap S^{n-1})} \geq 2^{-n}, \tag{4.36}$$

where E_δ is the energy window, n is the number of states in the energy window, and S^{n-1} is the $(n-1)$-sphere to force the states to be normalized. See appendix C.3 for more discussion about the volume of self-dual cones. Here we have used the property that $\mathcal{P}$ is dense and spans the Hilbert space, which also defines the size of $\mathcal{H}_\mathcal{P}$.

2. $|\Psi\rangle$ passes through $\mathcal{H}_\mathcal{P}$ often, namely 2^n times,

where n is the number of states in the energy window. Again, we use an exponentially small tolerance, thus points that differ by an amount less than the tolerance are identified. This is easy to prove for the real subspace spanned by the energy eigenvectors. A typical state passes through this subspace when all magnitudes are real, either positive or negative. We, therefore, pass 2^n times through this subspace, hitting all combinations of signs of the energy eigenvectors. It is reasonable to assume that any state passes 2^n times through $\mathcal{H}_\mathcal{P}$ as well, because they same degrees of freedom.

Another way to see this, is to look at equation (4.38) and see that, for a reasonable distribution of a_{ij}, the state passes through $\mathcal{H}_\mathcal{P}$ at least the assumed number of times.

Combining this with the result of the last assumption allows us to estimate the probability that any typical state passes through the natural cone

$$P_2 = 1 - (1 - P_1)^{(2^n)} \simeq 1 - e^{-1} \approx 0.6321. \tag{4.37}$$

This rough estimate of the probability is an $O(N^0)$ number, but is significantly smaller then one. We are, therefore, not certain that all states pass through the natural cone.

This is a very conservative estimate. For example, looking at equation (4.38) the number of passes through $\mathcal{H}_\mathcal{P}$ may be much larger then 2^n, in which case P_2 goes to one. Moreover, if the natural cone is more rounded, see appendix C.3, than the ratio of volumes is given by a^{-n}, with $\sqrt{2} < a < 2$, in which case P_2 goes to one as well. We can effectively get $P_2 = 1$ with one additional assumption, even with this conservative estimate.

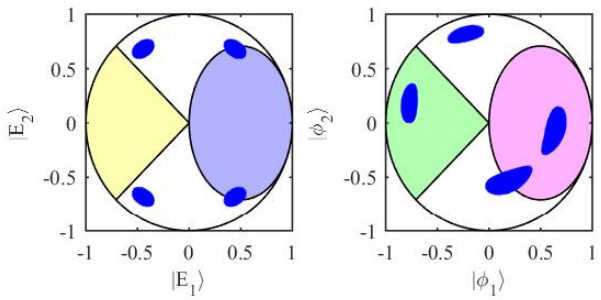


Figure 4.4: Left) Upper hemisphere is shown with the orthant (yellow) and circular (blue) cones. The blue dots are where a random test vector with time evolution gets close to the $\mathbb{R}^3$ subspace of $\mathbb{C}^3$, which is spanned by the energy eigenvalues. Right) The subspace is spanned by random vectors which causes the places where the test vector gets close the real subspace to shift and stretch. This shows that it is important that the natural cone is not contained in the real span of the energy eigenvectors.

3. Small perturbation around the state $|\Psi\rangle$ cause large changes in the time it takes to pass through $\mathcal{H}_{\mathcal{P}}$.

The phases need to be extremely fine tuned for a state to be in any specific real subspace. Thus while one state is in $\mathcal{H}_{\mathcal{P}}$, a different state that started close to the original state may need a lot of time before it passes through $\mathcal{H}_{\mathcal{P}}$ as well. Because of the large time differences a perturbed states passes through $\mathcal{H}_{\mathcal{P}}$ at points independent of the original state.

For example, consider a state with all coefficients real and positive at $t = 0$. It is, therefore, in the real subspace spanned by the energy eigenstates with real coefficients at time $t = 0$. Now consider a perturbation of this state, which gives a complex phase to a single energy eigenstate. This causes a large timeshift in the times when the perturbed state passes through the real subspace spanned by the energy eigenstates with real magnitudes. The points at which the perturbed states passes through the subspace is the same as the original state. This, however, is a consequence of the property that the energy eigenvalues form a real basis for the subspace and are the eigenstates of the Hamiltonian. This is not the case for $\mathcal{H}_{\mathcal{P}}$, and time evolution of a state will, therefore, change the magnitudes of the real basis vectors. We can write the time evolution of a typical state in the basis of the real subspace $\mathcal{H}_{\mathcal{P}}$ as follows

$$|\Psi(t)\rangle = \sum_i c_i e^{-iE_i t+\theta_i} |E_i\rangle = \sum_{i,j} a_{ij} e^{-iE_i t+\theta_{ij}} |j\rangle \,, \tag{4.38}$$

where we selected the coefficients c_i, a_{ij} to be real and positive and explicitly wrote the phases θ_i, θ_{ij}. The state lies on $\mathcal{H}_{\mathcal{P}}$ when the coefficient $\sum_i a_{ij} e^{-iE_i t+\theta_{ij}}$ in front of the basis vectors $|j\rangle$ is real for all basis vectors.

We can almost always find a state $|\Psi_1\rangle$ close to the original state $|\Psi_0\rangle = |\Psi_1\rangle + \frac{1}{N}|\delta\rangle$ that does pass through $\mathcal{P}$. The mirror operators $\widetilde{\mathcal{O}} = J\mathcal{O}J$ are bounded operators, the small perturbation from $|\delta\rangle$ will, therefore, only lead to subleading corrections. We can, therefore, use the mirror operators as fixed operators.

4.3.1.3 Some Subtleties

These results do not mean that we can apply the Tomita-Takesaki construction once and be done for all states. The requirement that the state is cyclic can exclude some states, but should be approximately true for most (almost all) typical states. It may seem that the restriction to the energy window is unnecessary, but this is not the case.

When applied to the black hole typical states we want the equation to be true at leading order, and this restricts the algebra $\mathcal{A}$. Most typical states cannot be annihilated by a small number of operators, small compared to the energy of the black hole. We, therefore, have a dependence on the energy window in the algebra.

Moreover, the modular Hamiltonian is not necessarily the same for all states in $\mathcal{P}$. Physics, however, tells us that we should identify the modular Hamiltonian as in equation (4.11) for most typical states. We again have a dependence on the energy window, in the form of the inverse temperature β. It can be expected that the various approximations cause subleading corrections, which makes correlators with both $\widetilde{\mathcal{O}}$ and $\mathcal{O}$ more sensitive, at the subleading level, to the microstate.

Time evolution does not change the cyclic and separating properties of states. Moreover, modular conjugation J is the same for all cyclic and separating states in $\mathcal{P}$. We can, thus, consider $\widetilde{\mathcal{O}} = JOJ$ to be a fixed operator when working on any typical state up to time evolution, as any state can almost be mapped into $\mathcal{P}$ by time evolution for a state-dependent amount. We will discuss the time-dependence in section 4.3.2.2.

Finally, there will be states that have strange time evolution compared to the typical states. Energy eigenstates only change by an overall phase and may be problematic for this construction. However, such states only constitute a small number of states in the energy window and are, therefore, not of interest to us. These caveats do not alter the conclusion that almost all state get close to $\mathcal{P}$ under time evolution.

4.3.2 Application of the Fixed Mirror Operators

In this subsection, we discuss how we can use the mirror operators in correlators.

4.3.2.1 How to Apply

Let us reiterate the rules we used when we applied the mirror operators in correlation functions.

$$\begin{aligned} \widetilde{\mathcal{O}}\,|\Psi_0\rangle &= e^{-\frac{\beta H}{2}}\mathcal{O}^\dagger e^{\frac{\beta H}{2}}\,|\Psi_0\rangle\,, \\ \widetilde{\mathcal{O}}A\,|\Psi_0\rangle &= A\widetilde{\mathcal{O}}\,|\Psi_0\rangle\,, \\ [H,\widetilde{\mathcal{O}}]A\,|\Psi_0\rangle &= Ae^{-\frac{\beta H}{2}}[H,\mathcal{O}^\dagger]e^{\frac{\beta H}{2}}\,|\Psi_0\rangle\,, \end{aligned} \tag{4.39}$$

where we left out the time arguments, as we will discuss them in the next subsection. These rules must be the same to leading order when we use a different typical state in the natural cone as reference state, for example,

$$|\Psi_1\rangle = A\widetilde{A}\,|\Psi_0\rangle\,. \tag{4.40}$$

The second is satisfied for $|\Psi_1\rangle$, because the modular conjugation J is the same in the natural cone. The third line is satisfied if the first line holds, because we identified the

modular Hamiltonian for typical states as in equation (4.11). So we only need to check the consistency of the first line. Using $|\Psi_0\rangle$ as the reference state we have

$$\langle\Psi_1| B_1\widetilde{B}_2 |\Psi_1\rangle = \langle\Psi_0| A^\dagger B_1 A e^{-\frac{\beta H}{2}}(A^\dagger B_2 A)^\dagger e^{\frac{\beta H}{2}} |\Psi_0\rangle\,, \tag{4.41}$$

while consistency of the first line of (4.39) demands that this is close to using $|\Psi_1\rangle = A\widetilde{A}|\Psi_0\rangle$ as the reference state

$$\langle\Psi_1| B_1\widetilde{B}_2 |\Psi_1\rangle = \langle\Psi_1| B_1 e^{-\frac{\beta H}{2}} B_2^\dagger e^{\frac{\beta H}{2}} |\Psi_1\rangle = \langle\Psi_0| A^\dagger B_1 e^{-\frac{\beta H}{2}} B_2^\dagger e^{\frac{\beta H}{2}} A e^{-\frac{\beta H}{2}} A^\dagger A e^{\frac{\beta H}{2}} |\Psi_0\rangle\,. \tag{4.42}$$

These two are, certainly, not equal for all A. However, A is restricted for typical states as discussed in section 4.3.1.1.

A Simple Example

We are interested in the case where A is constructed from Fourier modes with narrow spread, as shown in equation (4.34). We will work out what happens for a single Fourier mode and generalize from there. It is useful to work with the following modes

$$\begin{aligned}\langle a^\dagger a\rangle &= \frac{1}{e^{\beta\omega}-1}\delta(\omega-\omega') = g(\omega)\delta(\omega-\omega'),\\ \langle aa^\dagger\rangle &= \frac{e^{\beta\omega}}{e^{\beta\omega}-1}\delta(\omega-\omega') = e^{\beta\omega}g(\omega)\delta(\omega-\omega') = g(-\omega)\delta(\omega-\omega'),\end{aligned} \tag{4.43}$$

and consider $A = a$ up to normalization, which we can work out as follows

$$\langle\Psi_0| a^\dagger\widetilde{a}^\dagger\widetilde{a}a |\Psi_0\rangle = g(\omega)(1+2g(\omega))\delta^2, \tag{4.44}$$

where we omitted the arguments of the delta functions. Detailed calculations can be found in appendix C.1. The delta functions disappear when we integrate over the smearing of Fourier modes. We have to integrate four times, while we have two delta functions. We, therefore, obtain an expression for the normalization of the state $|\Psi_1\rangle = a\widetilde{a}/c_1|\Psi_0\rangle$

$$c_1 = \omega_s\sqrt{g(\omega)(1+2g(\omega))}, \tag{4.45}$$

where ω_s is the width of the smearing. We can now try to compare equations (4.41) and (4.42) with this preliminary work done. For this example we will use that the operator B_2 is some integral over Fourier modes $\int_0^{\omega^*} d\omega_2 K(\omega_2) b_{\omega_2}$.

$$\begin{aligned}\langle\Psi_0| A^\dagger B_1 A e^{-\frac{\beta H}{2}}(A^\dagger B_2 A)^\dagger e^{\frac{\beta H}{2}} |\Psi_0\rangle &= \int_0^{\omega^*} d\omega_2 K(\omega_2) c_1^{-2} \langle\Psi_0| a^\dagger B_1 a e^{-\frac{\beta H}{2}}(a^\dagger b_{\omega_2} a)^\dagger e^{\frac{\beta H}{2}} |\Psi_0\rangle\\ &= \int_0^{\omega^*} d\omega_2 K(\omega_2) c_1^{-2} e^{\frac{\omega_2\beta}{2}} \langle\Psi_0| a^\dagger B_1 a a^\dagger b_{\omega_2}^\dagger a) |\Psi_0\rangle\,,\end{aligned} \tag{4.46}$$

while the second line is

$$\begin{aligned}&\langle\Psi_0| A^\dagger B_1 e^{-\frac{\beta H}{2}} B_2^\dagger e^{\frac{\beta H}{2}} A e^{-\frac{\beta H}{2}} A^\dagger A e^{\frac{\beta H}{2}} |\Psi_0\rangle\\ &= \int_0^{\omega^*} d\omega_2 K(\omega_2) c_1^{-2} \langle\Psi_0| a^\dagger B_1 e^{-\frac{\beta H}{2}} b_{\omega_2}^\dagger e^{\frac{\beta H}{2}} a e^{-\frac{\beta H}{2}} a^\dagger a e^{\frac{\beta H}{2}} |\Psi_0\rangle\\ &= \int_0^{\omega^*} d\omega_2 K(\omega_2) c_1^{-2} e^{\frac{\omega_2\beta}{2}} \langle\Psi_0| a^\dagger B_1 b_{\omega_2}^\dagger a a^\dagger a |\Psi_0\rangle\,.\end{aligned} \tag{4.47}$$

We can, therefore, write the error we make, when we use the second line of equation (4.41) instead of the first line, as

$$\begin{aligned} \text{error} &= \int_0^{\omega^*} d\omega_2 K(\omega_2) c_1^{-2} e^{\frac{\omega_2 \beta}{2}} \langle \Psi_0 | a^\dagger B_1 [b_{\omega_2}^\dagger, a a^\dagger] a | \Psi_0 \rangle \\ &= - \int_0^{\omega^*} d\omega_2 K(\omega_2) c_1^{-2} e^{\frac{\omega_2 \beta}{2}} \langle \Psi_0 | a^\dagger B_1 a^\dagger a | \Psi_0 \rangle \, \delta(\omega - \omega_2), \end{aligned} \tag{4.48}$$

where we assumed that b and a are Fourier modes of the same operator and, therefore, obey canonical commutation relations. The constants c_1 in front are not important as we need to compare the error term to the size of equation (4.41) and that will have the same factor. The delta function limits the integration over the correlator with the integration kernel to the frequency window given by ω_s. Assuming that the kernel is smooth at ω, we see that the error term has a factor of $K(\omega)\omega_s$ relative to the correlators that we are comparing the error to.

We see that $|\Psi_1\rangle$ is no longer a typical state, as the state does not have a thermal distribution for the mode around ω. This does not change expectation values for operators in position space, because they are not sensitive to disturbances to single Fourier modes. Moreover, summing over many different disruptions, as in equation (4.32), should average out and return to an approximately thermal state, i.e. a typical state.

Generalization

We can generalize the previous example by considering $A = b_{\omega_1} b_{\omega_2} b_{\omega_3} ....$ Each frequency will generate a term with a delta function at that frequency. The sum over these cannot grow too large as the phases are usually random. The sum, therefore, only scales as $\sqrt{n}$, where n is the number of operators used. The maximum number of operators that we can use is also limited, as discussed in section 4.3.1.1, which ensures that the error will remain small.

The second generalization to consider are positive superpositions of these states. The same argument applies here. When we use a superposition of n states, we would have n^2 correlators. The phases, however, are random and the sum of these correlators only scale as n, which cancels with the normalization.

Working with more complicated operators B_1, B_2 also increases the number of error terms. We, however, consider simple operators, which have a bulk interpretation, and these operators are simple enough not to cause problems. The growth of the error is expected as more complicated operators are more sensitive to $O(1/N)$ corrections.

Perturbing the state a little bit with $|\delta\rangle$ introduces more error terms, see equation (4.32). The norm of this perturbation is, however, limited in size, which keeps these error terms under control.

These generalizations are enough for the cases that we are interested in. It is important to note that we clearly see subleading correction in the case that we have both normal and mirror operators in a correlator, while in the other cases the corrections are exponentially small, see section 4.2.1.2. We conclude that we can apply equation (4.39) also to $|\Psi_1\rangle$ at leading order, while the subleading corrections are different.

4.3.2.2 Time-Dependence of the Mirror Operators

We discussed the mirror operators in position space in section 4.1.3. We recall two of the equations that are useful for the discussion at this point. The Fourier transform

$$\widetilde{\mathcal{O}}_T(t) = \int_{-\omega_*}^{\omega_*} d\omega \; e^{-i\omega(t-T)} \widetilde{\mathcal{O}}_\omega, \tag{4.49}$$

shows two key aspects of the explicit time-dependence. First, the sign in front of the physical time t has the opposite sign from what we are used to in the Fourier transform. This means that the mirror operators go against the Killing isometry in the bulk. Secondly, there is the freedom of choosing T, which determines how much time evolution the operator has relative to the physical time. In position space we can approximately write

$$\widetilde{\mathcal{O}}_T(t) \left|\Psi_0\right\rangle = e^{-\frac{\beta H}{2}} \mathcal{O}^\dagger(T-t) e^{\frac{\beta H}{2}} \left|\Psi_0\right\rangle, \tag{4.50}$$

where we clearly see the sign flip of t and the time-shift by T. The other application rules deal with commutation relations, which are not affected by time dependence. This is enough to detail the time dependence of the operators when the mirror operators are defined for a single state.

We need to be more careful to get a consistent picture when we try to define the mirror operators for most typical states. It is natural to set $T = 0$ for the reference state $|\Psi_0\rangle$. This also sets $T = 0$ for typical states in the natural cone $\mathcal{P}$. For other typical states we use equation (4.32) to find the amount of time evolution that we need to get close to the natural cone, and use T to compensate that time evolution.

When doing an experiment with a typical state at some later time $|\Psi_1(t)\rangle$, we can always recover which mirror operator is highly entangled with by setting $T = 2t$ and work with the precursor of the mirror operator for that experiment. For example,

$$\left\langle\Psi_1(t_0)\right| \widetilde{\mathcal{O}}_{2t_0}(t_0+t')\mathcal{O}(t_0+t'') \left|\Psi_1(t_0)\right\rangle = \left\langle\Psi_1(t_0)\right| \mathcal{O}(t_0+t'') e^{-\frac{\beta H}{2}} \mathcal{O}^\dagger(t_0-t') e^{\frac{\beta H}{2}} \left|\Psi_1(t_0)\right\rangle, \tag{4.51}$$

this correlator is a function of $(t''+t')$ and is independent of t_0, and shows the entanglement structure at time t_0. The freedom to use the precursors is enhanced, because there is no reason to assume that the complexity of the mirror operators is lower at specific times.

It is important that we use a consistent time-shift T when we work with the precursors. We lose a consistent bulk interpretation otherwise. It is worth noting that all state-dependence is captured in the selection of T, i.e. if we want to construct a mirror operator with maximal entanglement at that time for a specific state, we find T and use it with the mirror operators from the reference state to generate the mirror operators for that specific state.

The modular conjugation operator J is approximately constant in time. This can be seen for the following property $J\Delta J = \Delta^{-1}$. This together with the antilinear nature of J means that $J\Delta^{is}J = \Delta^{is}$. However, time evolution is only approximately equal to Δ^{is}, which introduces a small amount of time evolution in the modular conjugation operator J and, therefore, in the definition of the Fourier modes of the mirror operator. Time evolution of J also introduces a time dependent phase. This phase, however, is not relevant as the correlation functions we consider always use an even number of J's.

Choice of Reference State

Most of the discussion so far made use of a reference state $|\Psi_0\rangle$ in their definitions, initially for the mirror operators and later for the natural cone $\mathcal{P}$. This, however, seems contrary to the claim that this construction is state-independent, except for the explicit time dependence. We already know that the mirror operators are the same for a typical state from the natural cone of the reference state, as $|\Psi_1\rangle \in \mathcal{P}_{|\Psi_0\rangle} \Rightarrow \mathcal{P}_{|\Psi_1\rangle} = \mathcal{P}_{|\Psi_0\rangle}$, provided that $|\Psi_1\rangle$ is cyclic.

We limit the discussion to the moment in time that a typical state is close to the natural cone for other typical states, as explicit time evolution will generate the answers for the other states. The leading results are the same for the reference state and the other typical state. Subleading results, however, may differ. The subleading corrections are, therefore, relative to the reference state and will change when we select a different reference state.

Another object that depends on the reference state is the small Hilbert space. The natural cone is part of the small Hilbert space, $\mathcal{P}_{|\Psi_0\rangle} \subset \mathcal{H}_{|\Psi_0\rangle}$, even though the product $A\widetilde{A}$ is not part of the small algebra. This happens because we can convert the mirror into normal operators as it is next to the state.

For any other typical state, at the moment that they are close to the natural cone, we have

$$\mathcal{H}_{(|\Psi_1\rangle+|\delta\rangle)} = \mathcal{H}_{|\Psi_0\rangle}, \tag{4.52}$$

moreover, time evolution does not alter the Hilbert space structure, so this holds at all times.

4.3.3 Consistency Checks

The construction is consistent when comparing different typical states. There are, however, other consistent checks that we need to do.

4.3.3.1 Superpositions

Superpositions are fundamental in quantum mechanics and our construction should be consistent when taking superpositions.

We can always take positive superpositions in the natural cone and remain in the natural cone because of the conical structure. This extends to superpositions of typical states that are at the same time, if we have defined $t = 0$ as the moment that they are close to the natural cone.

Other superposition, either not strictly positive or of typical states at different times, are problematic. These superpositions can result in a state that has a different time-shift compared to the states it is a superposition of. This means that the bulk interpretation the superposition can be very different from the bulk interpretations of the states that made up the superposition. This is different from the exterior, where we can interpret the superposition of states that have a particle in the exterior $\mathcal{O}(x,t)c_1 |\Psi_1\rangle + \mathcal{O}(x,t)c_2 |\Psi_2\rangle = \mathcal{O}(x,t) |\Psi_3\rangle$ as being a state having a particle in the same position. The interior is dependent on the time-shift and when the particle appears is not necessarily at the same time as before the superposition.

It is also possible to construct states that never get close to the natural cone and the interior operators cannot be defined for those states. This may seem rather destructive for

the construction, but it is exactly what we expect. We can create any state by taking arbitrary superpositions, including states with aberrant behavior. It is, therefore, consistent that we cannot take arbitrary superpositions, while keeping the interior operators fixed.

4.3.3.2 Perturbations

A critique of the original mirror operators [66] was that a perturbed state of the form $|\Psi_1\rangle = e^{ig\widetilde{\mathcal{O}}_\omega\widetilde{\mathcal{O}}_\omega^\dagger}|\Psi_0\rangle$ is also an equilibrium state and that we can define mirror operators for both $|\Psi_1\rangle$ and $|\Psi_0\rangle$, and conclude that both have smooth horizons. That both states have smooth horizons is in itself is not inconsistent. We must be careful when look at a state of the following form

$$|\Psi(t)\rangle = \begin{cases} |\Psi_0(t)\rangle & t < 0, \\ |\Psi_1(t)\rangle & t > 0. \end{cases} \tag{4.53}$$

Here we notice that it takes a time-shift to get $|\Psi_1\rangle$ close to the natural cone, as $e^{ig\widetilde{\mathcal{O}}_\omega\widetilde{\mathcal{O}}_\omega^\dagger}|\Psi_0\rangle$ is not in the natural cone. Explicit time dependence reduces correlation between mirror operators and normal operators after the perturbation, with a firewall as a consequence as expected by [66].

Conclusions

In this chapter we discussed whether explicit time dependence is enough to avoid the firewall paradoxes. This was inspired by the state-dependent mirror operators and the property that ergodic systems pass through all states of a given energy under time evolution. We have shown that it is possible to avoid the paradoxes with explicit time dependence, even though typical black hole microstates do not pass through all other typical black hole microstates under time evolution.

We have proposed to use the natural cone to get a construction of the interior operators. We have provided evidence that the natural cone defines the modular operator J for most typical states, thereby reducing the state dependence to a single parameter. Moreover, we know what this parameter does as this parameter comes from time evolution. This reduced state-dependence, therefore, reduces the fine-tuning needed for a smooth horizon and allows easier explicit construction in toy models.

Alternatively, if the typical states are not all captured in the natural cone and the time evolution thereof, then the typical states can be decomposed in multiple natural cones, see figure 4.5 for a toy model example. Remember that the natural cones of different states are the same if one state is in the natural cone of the other. Thus, if a test state and its time evolution is not in the natural cone of a reference state, then the natural cone of the test state has no overlap with the natural cone of the reference state. We can, therefore, define the mirror operators piece wise on independent natural cones, provided that the number of independent natural cones is not too large.

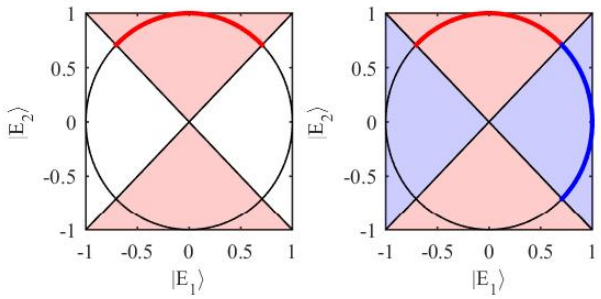


Figure 4.5: Left) In $\mathbb{C}^2$ it is possible to select a cone that only captures half of the states with time evolution. Right) We add a second cone to fill the space and are now able to capture all states.

Chapter 5

Concluding remarks

We asked the following question at the end of the introduction of this thesis: What do you experience when you fall into the black hole and cross the horizon? Nothing special, is the answer that we have gathered evidence for. However, each of the steps that we took to reach our conclusions is based on several assumption that, if proven wrong, could provide evidence for the opposite answer: Something special.

The first step that we took was applying the traversable wormhole protocol to the time-shifted thermofield double states. We concluded that the time-shifted thermofield double states are just as smooth as the thermofield double state. However, we saw that we needed to modify the protocol in a state-dependent manner to retrieve information send from one side of the wormhole to the other. The time-shifted thermofield double state are a idealized making typical states the natural direction for further exploration.

The second step was to apply the traversable wormhole protocol to typical states. We used the protocol to probe the geometry of typical state black holes and concluded that the geomtry should include part of the extended Schwarzschild geometry beyond the interior of the black hole. This region is described by the state-dependent mirror operators. The use of these mirror operators makes the traversable wormhole protocol state dependent for typical state black holes. We also concluded that the mirror operators must have explicit time dependence to avoid closed timelike curves.

Further applications of the traversable wormhole protocol and the mirror operators could be collapsing black holes in AdS and dS space. The AdS-vaidya geometry would be the next step in getting closer to black holes in our universe. Moreover, any realization of these construction in dS space would be a huge step forward for the same reason.

In the third step, we to further investigated explicit time dependence and the state dependence of the interior operators. We showed that explicit time dependence can be used to circumvent the firewall paradoxes. We also provided for a candidate construction for the interior operators in the form of the natural cone. These interior operators only have a single parameter state dependence, which is a great improvement over full state dependence. However, we did not prove that this construction is fully consistent and more work is needed in that regard. It would also be interesting to see how this proposal translates to the language of quantum information, which is being used more and more for the understanding of black holes.

The years leading up to this thesis have seen a lot of development in the study of black holes. Experimentally, the first black hole merges were observed using gravitational waves

and the first photograph of a black hole was taken. Several new tools were developed on the theory side of black holes. The soft hair proposal [67] should be mentioned in this regard. However, our focus on AdS made the SYK model and traversable wormhole protocol much more applicable. A large number of papers have either studied these tools or made use of them to study aspects of other systems. Only time will tell how the content of this thesis contributed to the larger understanding of black holes.

Appendix A

Appendix to chapter 2

A.1 Basic setup

We recall the basic setup used in [5, 9]. We start with the TFD state, inject a probe (representing the observer) in the left CFT at time $t = t_{\text{in}}$ by acting with $e^{i\epsilon\phi_L}$. We couple the two CFTs with $e^{ig\mathcal{O}_L\mathcal{O}_R}$ for a very short time around $t = 0$, and then calculate the expectation value of the outgoing probe on the right CFT with ϕ_R at $t = t_{\text{out}}$. All times are "laboratory time". We find it useful to think in terms of the wavefunction in the Schrödinger picture as a function of the laboratory time t. Ignoring the short amount of time that it takes to act with the operators mentioned above, the wavefunction is

$$|\Psi(t)\rangle = e^{-i(H_L+H_R)t}|\Psi_{\text{tfd}}\rangle \qquad t < t_{\text{in}} \tag{A.1}$$

$$|\Psi(t)\rangle = e^{-i(H_L+H_R)(t-t_{\text{in}})}e^{i\epsilon\phi_L}e^{-i(H_L+H_R)t_{\text{in}}}|\Psi_{\text{tfd}}\rangle \qquad t_{\text{in}} < t < 0 \tag{A.2}$$

$$|\Psi(t)\rangle = e^{-i(H_L+H_R)t}e^{ig\mathcal{O}_L\mathcal{O}_R}e^{i(H_L+H_R)t_{\text{in}}}e^{i\epsilon\phi_L}e^{-i(H_L+H_R)t_{\text{in}}}|\Psi_{\text{tfd}}\rangle \qquad 0 < t \tag{A.3}$$

Although the transitions are stated as sharp transitions, we should smear them a little to remove high energy modes. Next we compute the expectation value of ϕ_R at time $t = t_{\text{out}} > 0$:

$$\langle\Psi(t_{\text{out}})|\,\phi_R\,|\Psi(t_{\text{out}})\rangle = \langle\Psi_{\text{tfd}}|e^{i(H_L+H_R)t_{\text{in}}}e^{-i\epsilon\phi_L}e^{-i(H_L+H_R)t_{\text{in}}}e^{-ig\mathcal{O}_L\mathcal{O}_R}e^{i(H_L+H_R)t_{\text{out}}} \tag{A.4}$$

$$\times\,\phi_R e^{-i(H_L+H_R)t_{\text{out}}}e^{ig\mathcal{O}_L\mathcal{O}_R}e^{i(H_L+H_R)t_{\text{in}}}e^{i\epsilon\phi_L}e^{-i(H_L+H_R)t_{\text{in}}}|\Psi_{\text{tfd}}\rangle \tag{A.5}$$

$$= \langle\Psi_{\text{tfd}}|e^{-i\epsilon\phi_L(t_{\text{in}})}e^{-ig\mathcal{O}_L(0)\mathcal{O}_R(0)}\phi_R(t_{out})e^{ig\mathcal{O}_L(0)\mathcal{O}_R(0)}e^{i\epsilon\phi_L(t_{\text{in}})}|\Psi_{\text{tfd}}\rangle \tag{A.6}$$

Where we went from the Schrödinger picture to the Heisenberg picture to make the times more clear. This is the final correlator which can detect the excitation emerging in the right CFT after having traversed the wormhole.

In particular, by expanding the exponential $e^{i\epsilon\phi_L} \approx 1+i\epsilon\phi_L$, we obtain the commutator:

$$\langle[\phi_L(t_{\text{in}}), \phi_R(t_{\text{out}})]\rangle_V \neq 0 \tag{A.7}$$

Where V denotes the double-trace pertubation. This was shown by [9] to be nonzero, thereby demonstrating information transfer between the boundaries. Moreover, it was

argued that a multiparticle state is transfered without being destroyed, which is required for the path of the probe to be smooth. All this information is encoded in the correlator (A.4).

A.2 Time-shifted states

Let us now consider a time-shifted wormhole $|\Psi_T\rangle \equiv e^{iH_R T}|\Psi_{\text{tfd}}\rangle$ with $T > 0$. Note that, although we call this state a time-shifted state, we think of T as a parameter controlling the entanglement of the state. We would like to perform the same steps as before. We start with the TFD state, inject a probe in the left CFT at time $t = t_{\text{in}}$ by acting with $e^{i\epsilon\phi_L}$. As we will see, in order to make this state traversalbe, we now have to couple the two CFTs with $e^{ig\mathcal{O}_L X_R}$ for a very short time around $t = 0$, where X_R is a *different* operator than $\mathcal{O}_R$. The quantum state is

$$|\Psi_T(t)\rangle = e^{-i(H_L+H_R)t}e^{iH_R T}|\Psi_{\text{tfd}}\rangle \qquad t < t_{\text{in}} \tag{A.8}$$

$$|\Psi_T(t)\rangle = e^{-i(H_L+H_R)(t+t_w)}e^{i\epsilon\phi_L}e^{-i(H_L+H_R)t_{\text{in}}}e^{iH_R T}|\Psi_{\text{tfd}}\rangle \qquad t_{\text{in}} < t < 0 \tag{A.9}$$

$$|\Psi_T(t)\rangle = e^{-i(H_L+H_R)t}e^{ig\mathcal{O}_L X_R}e^{-i(H_L+H_R)t_w}e^{i\epsilon\phi_L}e^{i(H_L+H_R)t_w}e^{iH_R T}|\Psi_{\text{tfd}}\rangle \qquad 0 < t \tag{A.10}$$

We will now see that if we take X_R to be the precursor of the operator $O_R(T)$ then the quantum state of the right CFT at $t = T + t_{out}$ will be *exactly* the same as that of the TFD state at $t = t_{out}$. In particular it will contain the particle emerging out of the black hole in exactly the same form as in the TFD. Note that, because the right Hamiltonian commutes with left operators, we can rewrite the coupling between the boundary systems as follows $e^{ig\mathcal{O}_L X_R} = e^{ig\mathcal{O}_L e^{iH_R T}\mathcal{O}_R e^{-iH_R T}} = e^{iH_R T}e^{ig\mathcal{O}_L\mathcal{O}_R}e^{-iH_R T}$. We can diagonse this by computing the expectation value of operator ϕ_R on this state.

$$\langle\Psi_T(t_{\text{out}})|\,\phi_R\,|\Psi_T(t_{\text{out}})\rangle = \langle\Psi_{\text{tfd}}|e^{-iH_R T_1}e^{i(H_L+H_R)t_{\text{in}}}e^{-i\epsilon\phi_L}e^{-i(H_L+H_R)t_{\text{in}}} \tag{A.11}$$

$$\times e^{iH_R T_2}e^{-ig\mathcal{O}_L\mathcal{O}_R}e^{-iH_R T_2}e^{i(H_L+H_R)t_{\text{out}}}\phi_R e^{-i(H_L+H_R)t_{\text{out}}} \tag{A.12}$$

$$\times e^{iH_R T_2}e^{ig\mathcal{O}_L\mathcal{O}_R}e^{-iH_R T_2}e^{i(H_L+H_R)t_{\text{in}}}e^{i\epsilon\phi_L}e^{-i(H_L+H_R)t_{\text{in}}}e^{iH_R T_1}|\Psi_{\text{tfd}}\rangle \tag{A.13}$$

$$= \langle\Psi_{\text{tfd}}|e^{-iH_R(T_1-T_2)}e^{-i\epsilon\phi_L(t_{\text{in}})}e^{-ig\mathcal{O}_L(0)\mathcal{O}_R(0)}\phi_R(t_{\text{out}}-T_2) \tag{A.14}$$

$$\times e^{ig\mathcal{O}_L(0)\mathcal{O}_R(0)}e^{i\epsilon\phi_L(t_{\text{in}})}e^{iH_R(T_1-T_2)}|\Psi_{\text{tfd}}\rangle \tag{A.15}$$

For the case that $T_1 = T_2 = T$ we obtain:

$$\langle\Psi_T(t_{\text{out}})|\,\phi_R\,|\Psi_T(t_{\text{out}})\rangle_{V'} = \langle\Psi_0(t_{\text{out}}-T)|\,\phi_R\,|\Psi_0(t_{\text{out}}-T)\rangle_V \tag{A.16}$$

The time-shifted state is related to the normal state by a delay in the response. This is an *exact* statement: no approximations where made. The time-shifted state is just as traversable as the thermofield double state. However, the probe is received at a later time $t = t_{\text{out}} + T$. The probe does not feel this difference: the proper time of the proper does not

depend on T. The time shifted wormhole can, therefore, be used as a time-machine. By tuning the state and the corresponding perturbation, we can shift the probe by an arbitrary amount to the future, without changing the perception of the probe. An alternative to acting with the precursor on the right CFT, is acting with a precursor on the left CFT $e^{igY_L\mathcal{O}_R}$, or acting with precursors on both sides $e^{igY_LX_R}$. These options show the same results. Moreover, we can extend these results for $T < 0$ up to $T = -|t_{\text{ in}}| - t_{\text{ out}} + \epsilon$, where ϵ is some small number representing some limitations. For example, the sending of the probe, the scattering, and the receiving of the probe, are either smeared or nonlocal, and the precursors should not act within those areas. Therefore, we cannot receive the probe with zero time elapsed.

A.3 Additional shockwaves

It is necessary for the path of the probe to be smooth that the probe remembers what happened before encountering the shockwave, otherwise one may argue that the probe was killed earlier and regenerated by the shockwave. We consider the following state to study this:

$$|\Psi_T(t)\rangle = e^{-i(H_L+H_R)t}e^{iH_RT}|\Psi_{\text{ tfd}}\rangle \qquad t < t_{\text{ in}} \tag{A.17}$$

$$|\Psi_T(t)\rangle = e^{-i(H_L+H_R)(t-t_{\text{ in}})}e^{i\epsilon\phi_L}e^{-i(H_L+H_R)t_{\text{ in}}}e^{iH_RT}|\Psi_{\text{ tfd}}\rangle \qquad t_{\text{ in}} < t < t_s \tag{A.18}$$

$$|\Psi_T(t)\rangle = e^{-i(H_L+H_R)(t-t_s)}e^{ia\varphi_R}e^{-i(H_L+H_R)(t_s-t_{\text{ in}})}e^{i\epsilon\phi_L}e^{-i(H_L+H_R)t_{\text{ in}}}e^{iH_RT}|\Psi_{\text{ tfd}}\rangle \qquad t_s < t < 0 \tag{A.19}$$

$$|\Psi_T(t)\rangle = e^{-i(H_L+H_R)t}e^{ig\mathcal{O}_LX_R}e^{i(H_L+H_R)t_s}e^{ia\varphi_R} \tag{A.20}$$

$$\times\, e^{-i(H_L+H_R)(t_s-t_{\text{ in}})}e^{i\epsilon\phi_L}e^{-i(H_L+H_R)t_{\text{ in}}}e^{iH_RT}|\Psi_{\text{ tfd}}\rangle \qquad 0 < t \tag{A.21}$$

Where $e^{ia\varphi_R}$ generates a shockwave coming from the right, at some early time t_s. It is important that this shockwave is extremely weak, otherwise it will kill the probe. The double trace perturbation should still be able to extract the probe after the interaction of the additional shockwave. Moreover we want to use the same double trace perturbation, whether we would send this additional shockwave or not. An alternative would be to modify the double trace perturbation to remove this additional shockwave. That setup, however, cannot answer whether the memories of the probe are genuine. We rewrite the state, for $t > 0$, in such a way that it is clear that it corresponds to the same experiment in the thermofield state. The correlator follows directly from that.

$$|\Psi_T(t)\rangle = e^{-i(H_L+H_R)t}e^{ig\mathcal{O}_LX_R}e^{i(H_L+H_R)(t_s)}e^{ia\varphi_R}e^{-i(H_L+H_R)(t_s-t_{\text{ in}})}e^{i\epsilon\phi_L}e^{-i(H_L+H_R)t_{\text{ in}}}e^{iH_RT}|\Psi_{\text{ tfd}}\rangle \tag{A.22}$$

$$= e^{-i(H_L+H_R)(t-T)}e^{ig\mathcal{O}_L\mathcal{O}_R}e^{i(H_L+H_R)(t_s-T)}e^{ia\varphi_R}e^{-i(H_L+H_R)((t_s-T)-t_{\text{ in}})}e^{i\epsilon\phi_L}e^{-i(H_L+H_R)t_{\text{ in}}}|\Psi_{\text{ tfd}}\rangle \tag{A.23}$$

If we calculate $\langle\phi_R\rangle$ in this state we get the same response as in the thermofield state, with both the response and the additional shockwave being shifted by T. Thus we could extract information about the additional shockwave from the response.

Appendix B

Appendix to Chapter 3

B.1 The exterior geometry of typical black hole microstates

Here we briefly review some arguments which show that at large N the exterior geometry dual to a typical microstate should be the AdS-Schwarzschild geometry. The most conservative version of this statement is that, to leading order at large N and for time separations that are not too large, the boundary two-point function of light single trace operators on a typical pure state is close to the two-point function that would be computed from analytic continuation starting with the geometry of the Euclidean AdS black hole. By large N factorization this also implies that the leading large N disconnected part of higher point functions is also the same between the typical pure state and the eternal black hole.

We will start with the assumption that

$$Z^{-1}\ \mathrm{Tr}[e^{-\beta H}\mathcal{O}(\tau,\vec{x})\mathcal{O}(0,\vec{0})] = \langle\phi(\tau,\vec{x})\phi(0,\vec{0})\rangle_{\mathrm{EBH}} + O(1/N). \tag{B.1}$$

where the RHS is the boundary-limit of the two-point function of the dual bulk field that would be computed by a Witten diagram on the Euclidean black hole geometry (the subscript EBH refers to Euclidean Black Hole). We can not prove this statement, but we will take it as being true given that it is a basic prediction of the AdS/CFT correspondence at large N.

The first step is to analytically continue both sides to real time. The leading term on the RHS decays exponentially in time. For time scales of the order of scrambling time and longer, the leading term becomes comparable to the subleading term and we will not be able to control the approximations. Hence we restrict to time scales which are smaller than that. We can also restrict the correlators by considering only frequencies which do not scale with N, similar to the $\omega < \omega_*$ approximation used in the main text. For short time scales, and filtering out high frequencies, we have that the subleading $1/N$ corrections in the Euclidean computation will remain small. Hence we conclude that

$$Z^{-1}\ \mathrm{Tr}[e^{-\beta H}\mathcal{O}(t,\vec{x})\mathcal{O}(0,\vec{0})] = \langle\phi(t,\vec{x})\phi(0,\vec{0})\rangle_{\mathrm{EBH}} + O(1/N) \qquad , \qquad t \ll \beta\log N \tag{B.2}$$

The next step is to replace the canonical with the microcanonical ensemble. Since the RHS depends on the energy of the black hole only via the temperature, we expect that the saddle point approximation will be reliable and we conclude that

$$\mathrm{Tr}[\rho_m\mathcal{O}(t,\vec{x})\mathcal{O}(0,\vec{0})] = \langle\phi(t,\vec{x})\phi(0,\vec{0})\rangle_{\mathrm{EBH}} + O(1/N) \qquad , \qquad t \ll \beta\log N \tag{B.3}$$

The next step is to replace the microcanonical with the typical pure state. Here we expect the approximation of the LHS to be good up to exponentially small corrections, as discussed in section 3.4. Hence we find

$$\langle\Psi|\mathcal{O}(t,\vec{x})\mathcal{O}(0,\vec{0})|\Psi\rangle = \langle\phi(t,\vec{x})\phi(0,\vec{0})\rangle_{\text{EBH}} + O(1/N) \qquad , \qquad t \ll \beta \log N \tag{B.4}$$

as claimed.

In general we expect that the time separation on the boundary has an inverse relation to the distance from the horizon. The argument above suggests that there cannot be any modifications to the exterior geometry (or the state of the quantum field on top of the geometry) to any $O(N^0)$ distance from the horizon. The argument above does not exclude the possibility that there are modifications at —say— Planckian distance from the horizon, which would be detectable by an observer hovering very near the horizon.

However, for the infalling observer these possible modifications do not affect low point functions computed at macroscopic space and time separations in the frame of the infaller. For example, if we compute a two-point function between a point in the interior and a point in the exterior, separated by a distance of horizon scale, the two point function is robust even if we introduce cutoffs in the frequencies of the modes involved, as well as on the time scales over which we have access on the boundary [6, 68].

The argument above does not directly apply to the connected part of higher point functions, since that is suppressed by powers of $1/N$ and hence the perturbative expansion mixes up with the corrections of the order of $1/N$ from comparing the ensembles. It might be possible to disentangle the two types of corrections by taking the limit where the temperature of the black hole is very large. However, we will not explore this possibility further in this chapter.

Notice that similar approximations for time scales of the order of scrambling time, and for the $O(1)$ part of higher point functions, would follow from the conjecture of section 3.4 for this system, which remains currently unproven.

B.2 Time-dependence and Choice of T of the Mirror Operators

Since we are considering time-dependent perturbations, it is convenient to work in coordinate space, so we would like to define the mirror operators as a function of time. This suggests that we should Fourier transform $\widetilde{\mathcal{O}}_\omega$ back to coordinate time, modulo the limitations imposed by the restrictions on the frequencies discussed above. However here we encounter an interesting subtlety.

First of all we make a general observation in quantum mechanics: by specifying the matrix elements of an operator in the Heisenberg picture, as we have done in equations (4.12), it is not possible to conclude what is the physical time at which the given operator is localized. For instance, if we are given the matrix elements of two different operators, we do not have enough information to conclude what is the relative time-ordering of these two operators[1]. Hence, assigning a particular physical time to operators requires extra information that has to be specified in addition to the Heisenberg picture matrix elements[2].

[1] And to analyze time-dependent perturbations of the Hamiltonian it is necessary to have a well-defined notion of time-ordering.

[2] In the Schroedinger picture the localization of an operator in time is rather obvious.

For local operators $\mathcal{O}(t,x)$ in a CFT, it is natural to localize the operators in physical time which is given by the argument t. Notice however, that this is simply the most "natural" choice, but not a unique one. For example, we can imagine that the operator $\mathcal{O}(t,x)$ is localized at a time $t' \neq t$. This is the notion of the "precursor" of an operator [69].

In the case of local operators, the choice $t' = t$ is dictated by simplicity, as precursors are complicated operators. In the context of AdS/CFT the choice $t' = t$ for local operators, is also dictated by our desire to have a theory with a semiclassical bulk dual. If we consider a holographic CFT where the Hamiltonian is perturbed in a time-dependent fashion by local operators, for which the localization in physical time coincides with their argument t, i.e. if we avoid using perturbations by non-trivial precursors, then the dual geometry corresponds to solutions of gravity where the non-normalizable mode — corresponding to the source of the operator — is turned on in a particular time-dependent fashion. However if we consider perturbations by non-trivial precursors then the bulk interpretation may go beyond effective field theory, see for example [19] for a discussion.

This raises the question: how should we assign the mirror operators to any given physical time? In particular how should we select the time-ordering of the mirror operators? We will answer this question by imposing the following criterion: the mirror operators should be assigned a physical time in such a way that if we consider time-dependent perturbations of the Hamiltonian by mirror operators, the effect can be described by effective field theory in the bulk.

We already described in section 3.1.2.1 that according to this criterion there is a one-parameter family, labeled by T, of useful choices. All choices of the parameter T above lead to equivalent results, regarding the interpretation of the bulk geometry, if it is defined relationally with respect to the right boundary and if we do not turn on any sources for the $\widetilde{\mathcal{O}}$'s. However, since depending on the choice of T the mirror operators are anchored differently on physical time, the choices of T differ in what are the allowed time-dependent perturbations that we can do[3] — see example at the end of this subsection. For a given desired experiment, some choices of anchoring T may be more convenient than others.

We continue the discussion from section 3.1.2.1 to emphasize several non-trivial aspects of the time dependence of the mirror operators together with the choice of T. We write the (approximate) equations for the mirror operators in position space as follows,

$$\begin{aligned} \widetilde{\mathcal{O}}_T(t)\,|\Psi_0\rangle &= e^{-\frac{\beta H}{2}}\mathcal{O}^\dagger(T-t)e^{\frac{\beta H}{2}}\,|\Psi_0\rangle\,, \\ \widetilde{\mathcal{O}}_T(t)A(t_1,t_2,...)\,|\Psi_0\rangle &= A(t_1,t_2,...)\widetilde{\mathcal{O}}_T(t)\,|\Psi_0\rangle\,, \\ [H,\widetilde{\mathcal{O}}_T(t)]A(t_1,t_2,...)\,|\Psi_0\rangle &= A(t_1,t_2,..)e^{-\frac{\beta H}{2}}[H,\mathcal{O}^\dagger(T-t)]e^{\frac{\beta H}{2}}\,|\Psi_0\rangle\,. \end{aligned} \tag{B.5}$$

We emphasize again that we do not want to define the mirror operators for sharply time-localized operators $\mathcal{O}(t)$. The meaning of the equations above is that they are correct, provided that we convolute the operators with smearing functions in time, such that they pick out the Fourier modes with frequencies $|\omega| < \omega_*$. However, to simplify the notation we will continue writing $\widetilde{\mathcal{O}}(t)$, with the understanding that the operator has to be smeared out sufficiently. As we take ω_* to be larger, the approximation to a local operator becomes better.

[3] Of course we could also act with mirror operators which do not respect a single selection of T. This would not be fundamentally wrong, but it would not be consistent with a simple bulk dual geometry.

Comments

(1). Notice that the mirror operators as defined above obey

$$\widetilde{\mathcal{O}}_T(t) = \widetilde{\mathcal{O}}_{T'}(t - T + T'). \tag{B.6}$$

Hence for any choice of T, we are talking about the same set of operators, but anchored differently on physical time t. Equation (B.6) is an example of using "precursors" of operators, to localize them at different moments in time.
(2). In the bulk the choice of T corresponds to whether we think of the geometry as representing either the thermofield $|\Psi_{\text{tfd}}\rangle$ or one of its time-shifted cousins $e^{-iH_L T}|\Psi_{\text{tfd}}\rangle$, see [15]. Notice for instance

$$\langle\Psi|\mathcal{O}(t_1)\widetilde{\mathcal{O}}_T(t_2)|\Psi\rangle = \int d\omega_1 d\omega_2\, e^{-\frac{\beta\omega_2}{2}}\langle\Psi|\mathcal{O}_{\omega_1}\mathcal{O}^\dagger_{\omega_2}|\Psi\rangle e^{-i\omega_1 t_1 - i\omega_2(t_2-T)}, \tag{B.7}$$

given that for equilibrium states we have $\langle\Psi|\mathcal{O}_{\omega_1}\mathcal{O}^\dagger_{\omega_2}|\Psi\rangle \propto \delta(\omega_1 - \omega_2)$, we find the two-point function above is a function of $(t_1 + t_2 - T)$ and highly peaked around $t_1 + t_2 - T = 0$, which is the same as the behavior of two-sided correlators in the time-shifted thermofield states $e^{-iTH_L}|\text{TFD}\rangle$, where we take both times in the CFTs t_1, t_2 to run forward.
(3). We emphasize that if we only study correlators of operators in "autonomous states" in which the Hamiltonian is not time-dependent, then there is no way to distinguish between the different choices of the framing T.

Example

To illustrate the difference of possible choices of T, let us ask the physical question: what are the allowed perturbations of the form $\mathcal{O}\widetilde{\mathcal{O}}$ that we can perform at physical time $t = 0$. Suppose we select $T = 0$. Then as we see from the diagram we could perturb the Hamiltonian by

$$\delta H = \delta(t)\mathcal{O}(0)\widetilde{\mathcal{O}}_{T=0}(0). \tag{B.8}$$

The operators $\mathcal{O}(0)$ and $\widetilde{\mathcal{O}}_{T=0}(0)$ are highly entangled, and this perturbation would produce two shockwaves. However, if we make a choice of $T \neq 0$, and if we then ask what perturbation we are allowed to perform at $t = 0$ then we would have

$$\delta H = \delta(t)\mathcal{O}(0)\widetilde{\mathcal{O}}_T(0). \tag{B.9}$$

Now we notice from (B.7) that the operators $\mathcal{O}(0)$ and $\widetilde{\mathcal{O}}_T(0)$ are less entangled as $|T|$ increases, and the produced shockwave in the bulk becomes weaker.

Notice, however, that with a choice of $T \neq 0$, we can produce a strong shockwave by perturbing the state at time $t = T/2$ by

$$\delta H = \delta(t - T/2)\,\mathcal{O}(T/2)\,\widetilde{\mathcal{O}}_T(T/2). \tag{B.10}$$

since again $\mathcal{O}_T(T/2)$ and $\widetilde{\mathcal{O}}_T(T/2)$ from (B.7) are highly entangled.

B.3 Spherical shells on an Einstein-Rosen bridge

In this Appendix, we will study the gluing of spherical shells on an Einstein-Rosen bridge. The goal is to understand bulk energy associated to matter in the right and the left

regions. Using Einstein's equations, we will check that the mass of some matter to the left of the bifurcation horizon, if gravitationally dressed with respect to the right, is negative with respect to the right asymptotic boundary. This is similar to a particle constructed using mirror operator as discussed in Section 3.1.5.1 using purely CFT technology. The discussion here will be purely classical but it does show some aspects that are shared with the mirror operators. We will focus on 4d asymptotically flat space-time for ease of calculation. A useful reference is [70].

We will consider the initial value problem assuming we are at a moment of time symmetry. We parametrize the induced spacelike metric of the bridge as

$$ds^2 = \chi(\rho)^4(d\rho^2 + \rho^2 d\Omega_2^2). \tag{B.11}$$

For time-reflection invariant initial data (no extrinsic curvature) the Hamiltonian constraint of general relativity becomes $R = 0$, where R is the Ricci scalar of the spatial metric. For the ansatz (B.11) this becomes $\nabla^2\chi = 0$. The usual Einstein-Rosen bridge without matter is the solution

$$\chi = 1 + \frac{a}{\rho}. \tag{B.12}$$

ρ is not the same as the Schwarzchild radial coordinate r. Expanding near $\rho \to \infty$ we find $a = \frac{GM}{2}$. Moreover, we can multiply χ by an arbitrary constant without changing the physical metric. So more generally if we take $\chi = c_1 + \frac{c_2}{\rho}$ we find that the right mass is $M_R = \frac{2c_1c_2}{G}$. Notice that the left mass is also the same, which can be found by $\rho \to 1/\rho$.

We introduce a sphericall shell of matter which is momentarily at rest at $t = 0$, at a location $\rho = b$. This means that the solution will be

$$\chi = c_1 + \frac{c_2}{\rho}, \qquad \rho < b, \tag{B.13}$$

$$\chi = d_1 + \frac{d_2}{\rho}, \qquad \rho > b. \tag{B.14}$$

This depends on 5 parameters. One is trivial and eliminated by the overall rescaling. Another is eliminated by demanding continuity of χ at b. The others can be selected to correspond to: i) left mass M_L, ii) position of shell b iii) T_{00} of shell, which will be related to the "kink" of χ. These will fix M_R. Of course one can reorganize the independent parameters differently.

Continuity of the metric at the gluing demands that:

$$1 + \frac{GM_R}{2b} = c_1 + \frac{GM_L}{2c_1 b}. \tag{B.15}$$

The last parameter c_1 can be expressed in terms of the stress tensor of the shell of matter. For initial data with time reflection symmetry the constraint equation in the presence of matter reads $R = 16\pi G\rho_{\text{matter}}$, where $\rho_{\text{matter}} = T_{ab}n^a n^b$ and n^a is the unit normal vector. In this case the unit normal is $n^a = \frac{1}{\sqrt{f}}\delta^{a0}$, where $f = 1 - 2\frac{GM_R}{r}$, expressed in terms of ρ. We select an infinitely thin shell: $T_{00} = \mu f \delta(y - y_0)$, where y are locally flat radial coordinates. Converting to ρ this becomes

$$\rho_{\text{matter}} = \mu \frac{\delta(\rho - b)}{\chi^2|_{\rho=b}}. \tag{B.16}$$

We must, therefore, solve

$$-8\frac{\rho\chi'' + 2\chi'}{\rho\chi^5} = 16\pi G\rho_{\text{matter}} \tag{B.17}$$

to obtain c_1. This can be simplified across the shell.

$$\chi'(b^+) - \chi'(b^-) = -2\pi G\mu(\chi^3)_{\rho=b}, \tag{B.18}$$

$$-\frac{GM_R}{2b^2} + \frac{GM_L}{2c_1b^2} = -2\pi G\mu\left(1 + \frac{GM_R}{2b}\right)^3. \tag{B.19}$$

We can use this to eliminate c_1 from equation B.15 to obtain an expression for M_R

$$M_R = \frac{1 - 8\pi^2G^2b^2\mu^2 - \sqrt{1 + 4G^2\pi\mu M_L(1 - 2G\pi\mu b)}}{2\pi G^2\mu(2\pi Gb\mu - 1)}. \tag{B.20}$$

Expanding to linear order we find

$$M_R = M_L + 4\pi\mu(b^2 - \frac{G^2M_L^2}{4}) + O(\mu^2). \tag{B.21}$$

If we gravitationally dress the operators with respect to the right, then M_L is constant. Then we see that at linear order in μ the change of the mass M_R flips sign as we move the shell through the bifurcation point, which is at $b = \frac{GM_L}{2}$ before we add the shell. This is consistent with the commutation relations of the mirror operators with the CFT Hamiltonian in (4.12).

B.4 Numerics in the SYK model

The numerical study of the SYK model is straightforward for low enough values of N. The fermions satisfy the Clifford algebra

$$\{\psi_i, \psi_j\} = \delta_{i,j}\,, \tag{B.22}$$

where we have used a slightly different normalization than usual. This allows one to represent the fermions as Euclidean gamma matrices. For $N = 4$, for example, the matrices are,

$$\psi_1 = \frac{1}{\sqrt{2}}\begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}, \qquad \psi_2 = \frac{i}{\sqrt{2}}\begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{bmatrix},$$

$$\psi_3 = \frac{1}{\sqrt{2}}\begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{bmatrix}, \qquad \psi_4 = \frac{i}{\sqrt{2}}\begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{bmatrix}.$$

The Hamiltonian is the sum over q different products of these matrices, multiplied with a Gaussian-distributed random number, with mean zero. Calculations are fastest for the minimal q that still has interesting dynamics. Therefore, $q = 4$ is chosen,

$$H = \sum_{i<j<k<l} J_{ijkl}\psi_i\psi_j\psi_k\psi_l \qquad \text{Var}(J_{ijkl}) = \frac{6J^2}{N^3}, \tag{B.23}$$

where J sets the scale for the model. We can, therefore, set $J = 1$ to simplify this further. The energy levels and energy eigenvectors can be obtain by diagonalizing the Hamiltonian. The typical state is constructed as a superposition of these vectors.

Taking the usual Laplace transform to get the relation between energy and temperature does not work because the energy levels are not smoothed out for low N, see for example figure B.5. Instead, the temperature is fixed by demanding that the average energy in the typical pure state is the same as that in the thermal ensemble. This leads to an order $O(1/N)$ error for the temperature in the large N limit, but it is computationally fast. For the numerics in this chapter, $\beta = 5$ is chosen for the inverse temperature.

In the rest of this appendix we will show several figures that were used in the main text at different N. We compare the thermal correlators with the pure state correlators,

$$\langle A \rangle_{\text{Thermal}} = \frac{\text{Tr}[e^{-\beta H} A]}{\text{Tr}[e^{-\beta H}]} , \qquad \langle A \rangle_{\text{Pure}} = \langle \Psi_0 | A | \Psi_0 \rangle . \tag{B.24}$$

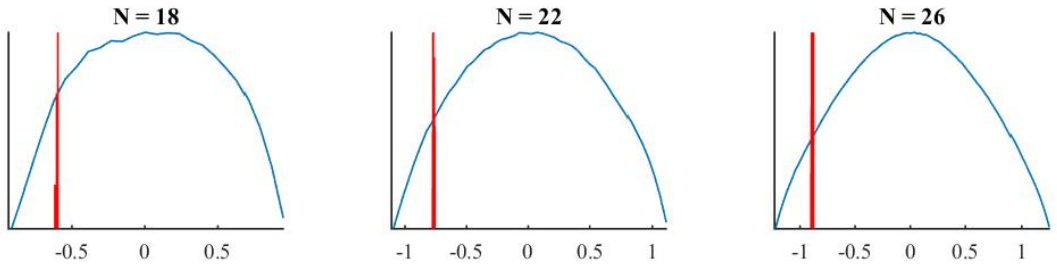


Figure B.1: Density of states of the SYK model (blue), and the energy eigenstates excited in the equilibrium state $|\Psi_0\rangle$ (red).

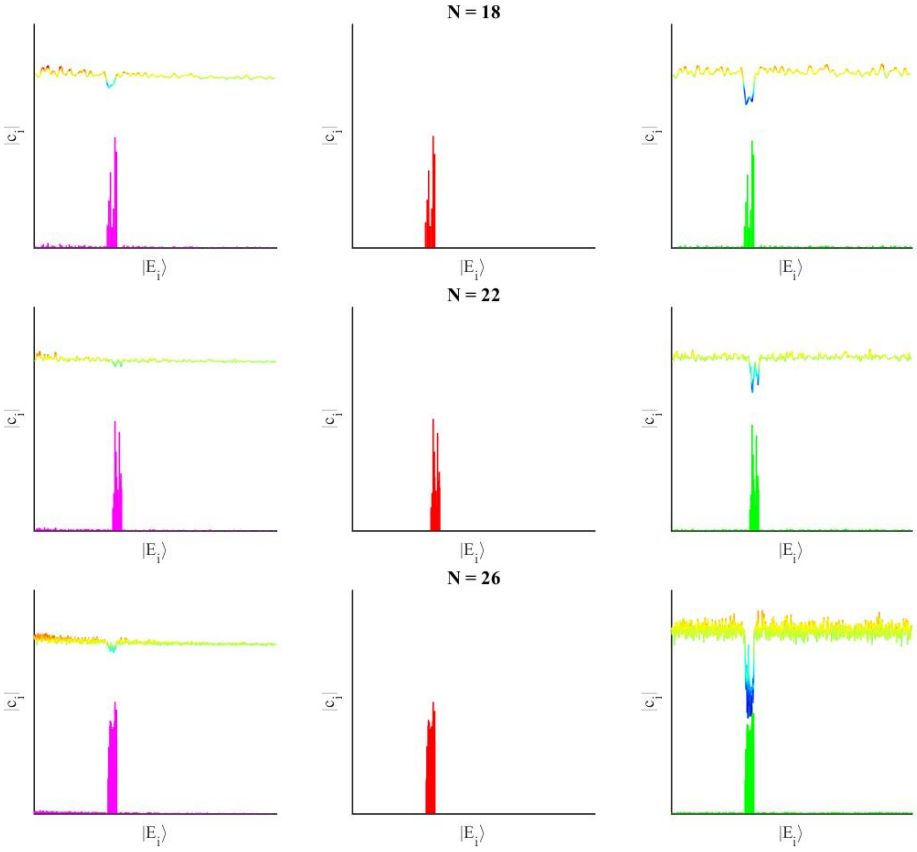


Figure B.2: Distribution of $|c_i|$ in non-equilibrium state of the form $U(\widetilde{S}_i)|\Psi_0\rangle$ (magenta), typical equilibrium state $|\Psi_0\rangle$ (red), and non-equilibrium state of the form $U(S_i)|\Psi_0\rangle$ (green). The line above the bar plot shows, in heat map colors, which eigenstates are excited, and which ones are suppressed because of the perturbation. Blue eigenstates are suppressed, while eigenstates with other colors are excited with small (green), medium (orange), or large (red) magnitude.

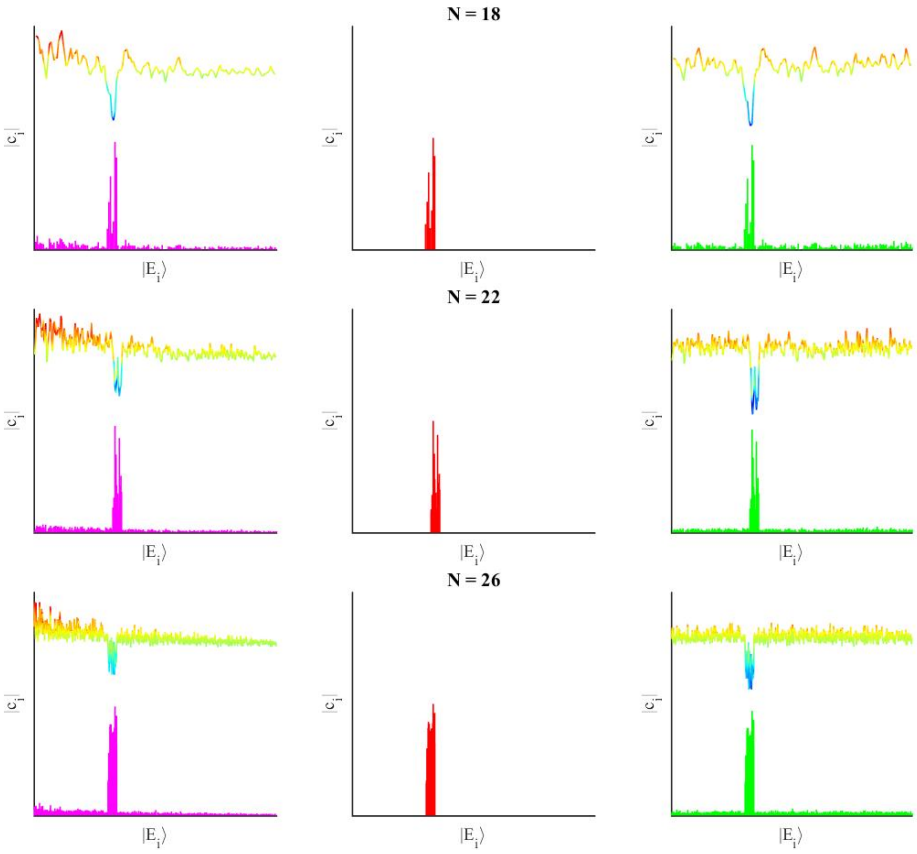


Figure B.3: Distribution of $|c_i|$ in non-equilibrium state of the form $U(\widetilde{S}_i)|\Psi_0\rangle$ (magenta), typical equilibrium state $|\Psi_0\rangle$ (red), and non-equilibrium state of the form $U(S_i)|\Psi_0\rangle$ (green). The line above the bar plot shows, in heat map colors, which eigenstates are excited, and which ones are suppressed because of the perturbation. Blue eigenstates are suppressed, while eigenstates with other colors are excited with small (green), medium (orange), or large (red) magnitude.

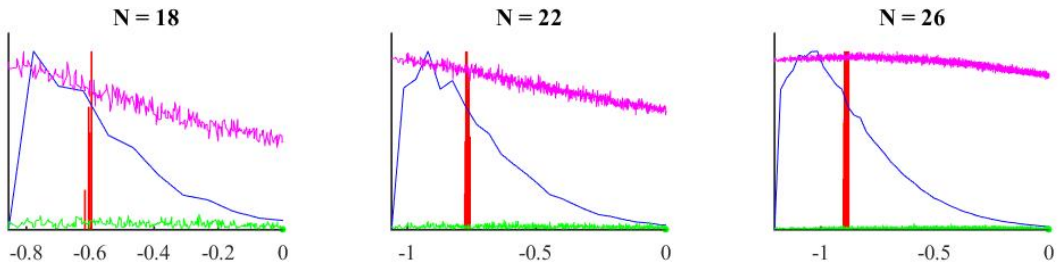


Figure B.4: The diagonal elements of $\{\psi_1(t), \psi_2(0)\}^2$ at scrambling time (magenta) dominate over the superdiagonal (green) and is slowly varying in the energy region of the thermal ensemble (blue) and the equilibrium state (red).

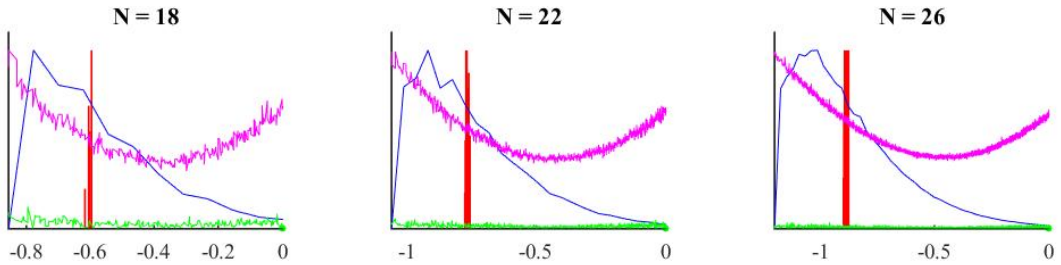


Figure B.5: The diagonal elements of $-[S_1(t), S_2(0)]^2$ at scrambling time (magenta) dominate over the superdiagonal (green) and is slowly varying in the energy region of the thermal ensemble (blue) and the equilibrium state (red).

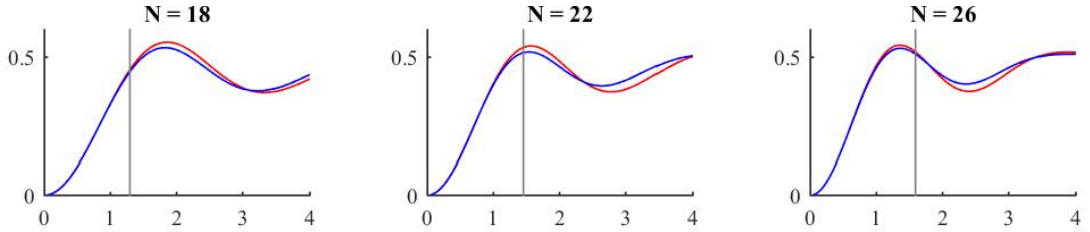


Figure B.6: We compare the expectation value of $\langle\{\psi_1(t), \psi_2(0)\}^2\rangle$ in the thermal state (blue) and pure state (red). The scrambling time is designated by the vertical line.

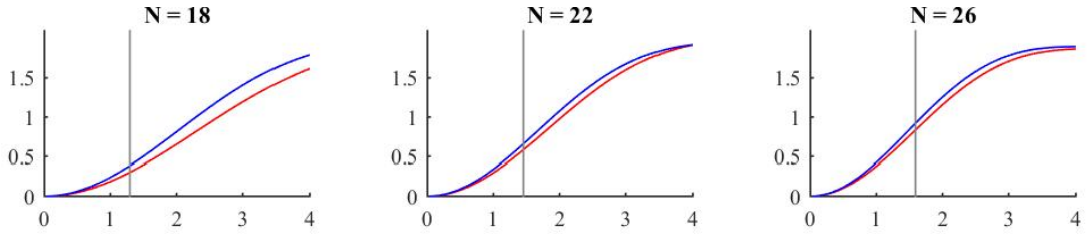


Figure B.7: We compare the expectation value of $-\langle[S_1(t), S_2(0)]^2\rangle$ in the thermal state (blue) and pure state (red). The scrambling time is designated by the vertical line.

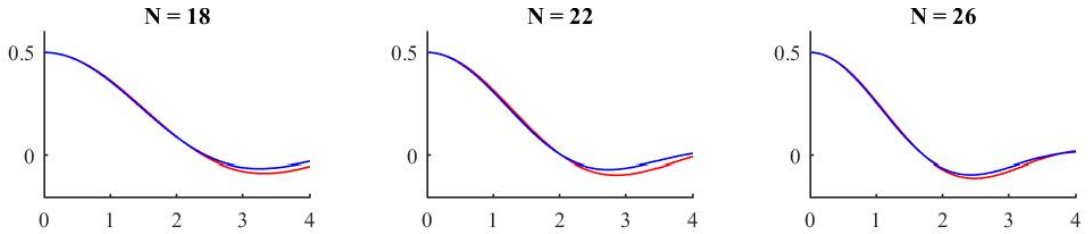


Figure B.8: We compare the expectation value of $\langle\psi_1(t)\psi_1(0)\rangle$ in the thermal state (blue) and pure state (red).

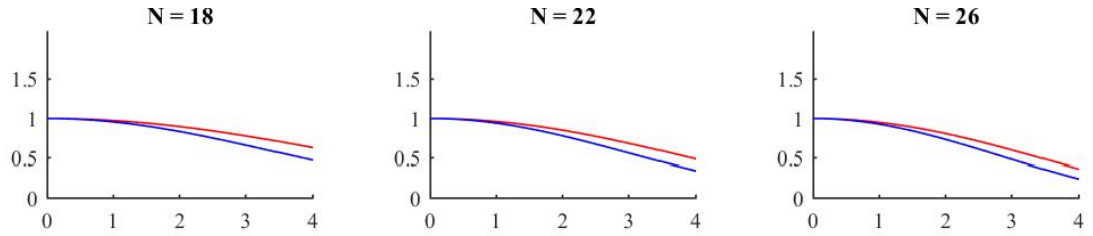


Figure B.9: We compare the expectation value of $\langle S_1(t)S_1(0)\rangle$ in the thermal state (blue) and pure state (red).

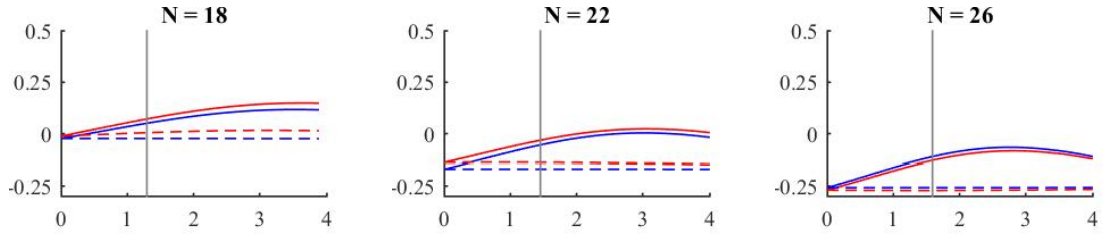


Figure B.10: Comparing the traversable wormhole correlator (3.92) in the thermal state (blue) and typical state (red). The dotted line is the signal without the probe, which is the response one obtains from just the state-dependent perturbation and should disappear in the large N limit, in this case a sum over five pairs of operators was used in the perturbation to limit this effect. The vertical line denotes the scrambling time and is the time around which the probe is focused.

Appendix C

Appendix to Chapter 4

C.1 Thermal correlators

In this appendix we investigate several correlators that are relevant to the discussions in section 4.3.1.1 and section 4.3.2. We will only focus on the leading order in all calculations in this appendix.

For a Fourier mode a at frequency ω we can write the commutation relations as

$$[a, a^\dagger] = \delta(\omega - \omega'). \tag{C.1}$$

Moreover, we can use the commutation relations with the Hamiltonian to obtain the following expression

$$e^{-\beta H} a = e^{\beta\omega} a e^{-\beta H}. \tag{C.2}$$

These two equations are enough to calculate the correlators we are interested in.

$$\begin{aligned}\langle a^\dagger a\rangle = &\ \mathrm{Tr}[e^{-\beta H} a^\dagger a]/Z = -\delta(\omega-\omega') + \ \mathrm{Tr}[e^{-\beta H} a a^\dagger]/Z = -\delta(\omega-\omega') + e^{\beta\omega}\ \mathrm{Tr}[e^{-\beta H} a^\dagger a]/Z \\ &= \frac{1}{e^{\beta\omega}-1}\delta(\omega-\omega'),\end{aligned} \tag{C.3}$$

where we used that the correlators in the micro-state are thermal to leading order. We obtain among similar lines the following

$$\langle a a^\dagger\rangle = \frac{e^{\beta\omega}}{e^{\beta\omega}-1}\delta(\omega-\omega'). \tag{C.4}$$

We can also obtain the two point function between a normal and a mirror operator in this manner

$$\langle a\widetilde{a}\rangle = \langle a e^{-\beta H/2} a^\dagger e^{\beta H/2}\rangle = e^{-\beta\omega/2}\langle a a^\dagger\rangle = \frac{e^{\beta\omega/2}}{e^{\beta\omega}-1}\delta(\omega-\omega'), \tag{C.5}$$

and similarly

$$\langle a^\dagger \widetilde{a}^\dagger\rangle = \frac{e^{\beta\omega/2}}{e^{\beta\omega}-1}\delta(\omega-\omega'). \tag{C.6}$$

We are also interested in the norm of a state of the form $|\Psi_1\rangle \sim \widetilde{a}a\,|\Psi_0\rangle$, where the Fourier modes are slightly smeared. We, therefore, calculate the following four point function

$$\begin{aligned}
\langle a^\dagger \widetilde{a}^\dagger \widetilde{a} a\rangle &= \langle a^\dagger a a^\dagger a\rangle \\
&= \mathrm{Tr}[e^{-\beta H} a^\dagger a a^\dagger a]/Z \\
&= \mathrm{Tr}[e^{-\beta H} a a^\dagger a^\dagger a]/Z - \mathrm{Tr}[e^{-\beta H} a^\dagger a]/Z\delta(\omega-\omega') \\
&= \mathrm{Tr}[e^{-\beta H} a a^\dagger a a^\dagger]/Z - \mathrm{Tr}[e^{-\beta H} a^\dagger a]/Z\delta(\omega-\omega') - \mathrm{Tr}[e^{-\beta H} a a^\dagger]/Z\delta(\omega''-\omega''') \\
&= e^{\beta\omega}\,\mathrm{Tr}[e^{-\beta H} a^\dagger a a^\dagger a]/Z - \mathrm{Tr}[e^{-\beta H} a^\dagger a]/Z\delta(\omega-\omega') - \mathrm{Tr}[e^{-\beta H} a a^\dagger]/Z\delta(\omega''-\omega''') \\
&= \frac{1}{1-e^{\beta\omega}}\Big(-\langle a^\dagger a\rangle\delta(\omega-\omega') - \langle a a^\dagger\rangle\delta(\omega''-\omega''')\Big) \\
&= \frac{e^{\beta\omega}+1}{(e^{\beta\omega}-1)^2}\delta(\omega-\omega')\delta(\omega''-\omega'''),
\end{aligned} \tag{C.7}$$

where we used that $a^\dagger a$ commutes with the Hamiltonian in the first line. We have not been precise which Fourier mode we exactly track by using $\omega = \omega''$. The error is suppressed when we consider the smearing $a = \int d\omega' a_{\omega'}$, with integration bounds $(\omega - \frac{1}{2}\omega_s, \omega + \frac{1}{2}\omega_s)$. We then obtain

$$\langle a^\dagger \widetilde{a}^\dagger \widetilde{a} a\rangle = \frac{e^{\beta\omega}+1}{(e^{\beta\omega}-1)^2}\omega_s^2, \tag{C.8}$$

and similarly

$$\langle a\widetilde{a}\widetilde{a}^\dagger a^\dagger\rangle = e^{\beta\omega}\frac{e^{\beta\omega}+1}{(e^{\beta\omega}-1)^2}\omega_s^2. \tag{C.9}$$

We can, therefore, compute the overlap between $|\Psi_0\rangle$ and $|\Psi_1\rangle = \widetilde{a}a\,|\Psi_0\rangle\,/\mathcal{N}$, where $\mathcal{N}$ is the normalization,

$$\langle\Psi_0|\Psi_1\rangle = \langle\widetilde{a}a\rangle/\sqrt{\frac{e^{\beta\omega}+1}{(e^{\beta\omega}-1)^2}\omega_s^2} = \frac{e^{\beta\omega/2}}{\sqrt{e^{\beta\omega}+1}}, \tag{C.10}$$

and similarly for the state $|\Psi_1'\rangle \sim \widetilde{a}^\dagger a^\dagger\,|\Psi_0\rangle$ we obtain the overlap

$$\langle\Psi_0|\Psi_1'\rangle = \frac{1}{\sqrt{e^{\beta\omega}+1}}. \tag{C.11}$$

We can also calculate the overlap between two states perturbed different frequencies, i.e. $|\Psi_1\rangle \sim \widetilde{a}a\,|\Psi_0\rangle$ and $|\Psi_2\rangle \sim \widetilde{b}b\,|\Psi_0\rangle$. The relevant four point function is

$$\begin{aligned}
\langle b^\dagger \widetilde{b}^\dagger \widetilde{a} a\rangle &= \langle b^\dagger a e^{-\beta H/2} a^\dagger b e^{\beta H/2}\rangle \\
&= e^{-\beta(\omega_a-\omega_b)/2}\langle b^\dagger a a^\dagger b\rangle \\
&= e^{-\beta(\omega_a-\omega_b)/2}\Big(\mathrm{Tr}[e^{-\beta H} b^\dagger b a a^\dagger]/Z\Big) \\
&= e^{-\beta(\omega_a-\omega_b)/2}\Big(\mathrm{Tr}[e^{-\beta H} b^\dagger b a^\dagger a]/Z + \mathrm{Tr}[e^{-\beta H} b^\dagger b]/Z\delta(\omega_a-\omega_a')\Big) \\
&= e^{-\beta(\omega_a-\omega_b)/2}\Big(e^{-\beta\omega_a}\,\mathrm{Tr}[e^{-\beta H} b^\dagger b a a^\dagger]/Z + \langle b^\dagger b\rangle\delta(\omega_a-\omega_a')\Big) \\
&= \frac{e^{-\beta(\omega_a-\omega_b)/2}}{(1-e^{-\beta\omega_a})(e^{\beta\omega_b}-1)}\delta(\omega_a-\omega_a')\delta(\omega_b-\omega_b') \\
&= \frac{e^{\beta(\omega_a+\omega_b)/2}}{(e^{\beta\omega_a}-1)(e^{\beta\omega_b}-1)}\delta(\omega_a-\omega_a')\delta(\omega_b-\omega_b'),
\end{aligned} \tag{C.12}$$

which we use to calculate the overlap, after taking the smearing into account

$$\begin{aligned}\langle\Psi_2|\Psi_1\rangle &= \frac{e^{\beta(\omega_a+\omega_b)/2}}{(e^{\beta\omega_a}-1)(e^{\beta\omega_b}-1)}\sqrt{\frac{(e^{\beta\omega_a}-1)^2}{e^{\beta\omega_a}+1}}\sqrt{\frac{(e^{\beta\omega_b}-1)^2}{e^{\beta\omega_b}+1}} \\ &= \frac{e^{\beta(\omega_a+\omega_b)/2}}{\sqrt{e^{\beta\omega_a}+1}\sqrt{e^{\beta\omega_b}+1}}.\end{aligned} \tag{C.13}$$

In the limit that the frequencies are almost the same, but still in different bins, we obtain

$$\langle\Psi_2|\Psi_1\rangle = \frac{e^{\beta\omega}}{e^{\beta\omega}+1}. \tag{C.14}$$

We can obtain the expression for the states $|\Psi_1'\rangle \sim \widetilde{a}^\dagger a^\dagger |\Psi_0\rangle$ and $|\Psi_2'\rangle \sim \widetilde{b}^\dagger b^\dagger |\Psi_0\rangle$ in the same limit in a similar manner

$$\langle\Psi_2'|\Psi_1'\rangle = \frac{1}{e^{\beta\omega}+1}. \tag{C.15}$$

To conclude this appendix, we note that the most important result is that the overlap is significantly not equal to one for a large range of frequencies. This is necessary for the assumption that the states are independent.

C.2 Overlap of states

In this appendix, we try to estimate how many states we need to get close to a given test state, recall equation (4.32),

$$|\Psi_1(t_0)\rangle + |\delta\rangle = a_0|\Psi_0\rangle + a_1A_1\widetilde{A}_1|\Psi_0\rangle + a_2A_2\widetilde{A}_2|\Psi_0\rangle ... \tag{C.16}$$

This question cannot be answered in general. Therefore, we study a simpler problem. Given a test state, how many random states do we need to get close to the test state.

$$|\Psi(t_0)\rangle + |\delta\rangle = a_1|\Psi_1\rangle + a_2|\Psi_2\rangle + a_3|\Psi_3\rangle ... \tag{C.17}$$

We can rewrite this as a least square problem, where the matrix A has the states $|\Psi_i\rangle$ as its columns and the vector x has the coefficients a_i as its entries.

$$Ax = |\Psi(t_0)\rangle, \tag{C.18}$$

and maximize over the phases after obtaining the least squares solution. Maximizing over the phases is a computational hard problem and we, therefore, rephrase the problem as follows. We instead maximize

$$M = \max_{t_0}(\langle\Psi(t_0)|\, Ax), \tag{C.19}$$

first over the phases, which we can do by taking the elementwise absolute value,

$$M = |\langle\Psi|| \cdot |Ax|, \tag{C.20}$$

and then numerically maximize over x. The results are shown in figure C.1. We use interpolation to get a clearer view of how the overlap develops.

These results suggest that there is linear relation between the length of the vectors and the number of vectors in the superposition for constant overlap. However, numerical algorithms can only find local extremes, and by repeating this many times an estimate for the global minimum is obtained. The results are, therefore, too small. This effect becomes stronger with a larger parameter space (more vectors in the superposition) and a more rapidly changing function (longer vectors).

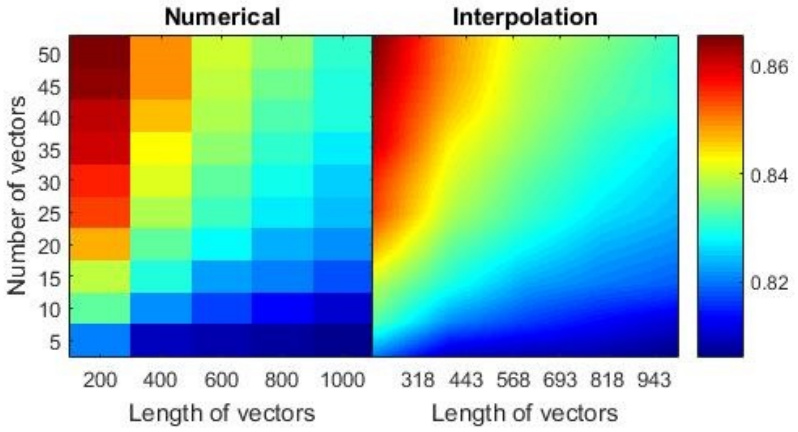


Figure C.1: Numerical search for the largest overlap following the method described in equation C.20.

C.3 Volume of self-dual cones

In this appendix, we will discuss some shapes of self dual cones and how they affect the discussion about the natural cones. We will restrict ourselves to Hilbert spaces of the form $\mathbb{R}^n$ to simplify notation. We can always restrict to the real subspace that a cone lies on in a complex Hilbert space. We repeat the definition of the dual of a cone as a reminder

$$K^D \coloneqq \{|x\rangle \in \mathcal{H} : \langle x|y\rangle \geq 0, \forall\, |y\rangle \in K\}. \tag{C.21}$$

The volume of the natural cone is difficult to estimate. However, the volume of the natural cone is constrained to between the orthant cone and the Lorentz cone, which are the two examples that we will discuss.

C.3.1 The Orthant Cone

The most straightforward cone is the orthant cone. This the cone that has non-negative amplitudes in some basis.

$$K_O \coloneqq \{|x\rangle \in \mathcal{H} \quad | \quad |x\rangle = \sum_i a_i \,|i\rangle\,, a_i \geq 0\}, \tag{C.22}$$

where $|i\rangle$ are the basis vectors. It is trivial to show that this is a cone, that it is convex, and that it is self-dual. We can generate a new cone by flipping the sign of one (or more) of the basis vectors, and cover the Hilbertspace $\mathcal{H}$ by doing this in all possible combinations, i.e. by 2^n orthant cones. This means that the volume of one of these cones is given by

$$\frac{\mathrm{Vol}(K_O \cap S^{n-1})}{\mathrm{Vol}(\mathcal{H} \cap S^{n-1})} = 2^{-n}, \tag{C.23}$$

where n is the dimension of the Hilbert space and S^{n-1} is the $(n-1)$-sphere to restrict to normalized states.

C.3.2 The Circular Cone

Another example of a commonly discussed cone is the circular cone.

$$K_{|u\rangle,\theta} \coloneqq \{|x\rangle \in \mathcal{H} \quad | \quad \measuredangle(|x\rangle, |u\rangle) \leq \theta\}, \tag{C.24}$$

for some axis of rotation $|u\rangle$ and angle θ. The dual of this cone is given by

$$K^D_{|u\rangle,\theta} = K_{|u\rangle,\frac{\pi}{2}-\theta}. \tag{C.25}$$

The cone is, therefore, self-dual for the angle $\theta = \pi/4$. This coincides with the orthant cone in two dimensions.

The self-dual circular cone is called the Lorentz cone.

$$K_L \coloneqq \{(|x\rangle, t) \in \mathcal{H} \quad | \quad ||x||_2 \leq t\}, \tag{C.26}$$

where we could use any basis vector for t, this direction is excluded from $|x\rangle$.

The volume of a circular cone is the area of a spherical cap, which for the case of $\theta = \pi/4$, is given by

$$\frac{\mathrm{Vol}(K_{|u\rangle,\pi/4} \cap S^{n-1})}{\mathrm{Vol}(\mathcal{H} \cap S^{n-1})} = \frac{1}{2} I_{1/2}\left(\frac{n-1}{2}, \frac{1}{2}\right) \propto n^{-1/2}(\sqrt{2})^{-n}, \tag{C.27}$$

where $I_x(a, b)$ is the regularized incomplete beta function.

Bibliography

[1] R. van Breukelen and K. Papadodimas, *Quantum teleportation through time-shifted AdS wormholes*, *JHEP* **08** (2018) 142 [1708.09370].

[2] J. de Boer, R. Van Breukelen, S. F. Lokhande, K. Papadodimas and E. Verlinde, *On the interior geometry of a typical black hole microstate*, *JHEP* **05** (2019) 010 [1804.10580].

[3] J. De Boer, R. Van Breukelen, S. F. Lokhande, K. Papadodimas and E. Verlinde, *Probing typical black hole microstates*, *JHEP* **01** (2020) 062 [1901.08527].

[4] R. van Breukelen, *Black Hole State Dependence as a Single Parameter*, *JHEP* **04** (2020) 210 [1910.00036].

[5] P. Gao, D. L. Jafferis and A. Wall, *Traversable Wormholes via a Double Trace Deformation*, 1608.05687.

[6] K. Papadodimas and S. Raju, *An Infalling Observer in AdS/CFT*, *JHEP* **1310** (2013) 212 [1211.6767].

[7] K. Papadodimas and S. Raju, *The Black Hole Interior in AdS/CFT and the Information Paradox*, 1310.6334.

[8] K. Papadodimas and S. Raju, *State-Dependent Bulk-Boundary Maps and Black Hole Complementarity*, 1310.6335.

[9] J. Maldacena, D. Stanford and Z. Yang, *Diving into traversable wormholes*, *Fortsch. Phys.* **65** (2017) 1700034 [1704.05333].

[10] J. Maldacena and L. Susskind, *Cool horizons for entangled black holes*, 1306.0533.

[11] A. Almheiri, D. Marolf, J. Polchinski and J. Sully, *Black Holes: Complementarity or Firewalls?*, *JHEP* **02** (2013) 062 [1207.3123].

[12] A. Almheiri, D. Marolf, J. Polchinski, D. Stanford and J. Sully, *An Apologia for Firewalls*, *JHEP* **1309** (2013) 018 [1304.6483].

[13] D. Marolf and J. Polchinski, *Gauge/Gravity Duality and the Black Hole Interior*, *Phys.Rev.Lett.* **111** (2013) 171301 [1307.4706].

[14] K. Papadodimas and S. Raju, *The unreasonable effectiveness of exponentially suppressed corrections in preserving information*, *Int. J. Mod. Phys.* **D22** (2013) 1342030.

[15] K. Papadodimas and S. Raju, *Local Operators in the Eternal Black Hole*, *Phys. Rev. Lett.* **115** (2015) 211601 [1502.06692].

[16] K. Papadodimas and S. Raju, *Remarks on the necessity and implications of state-dependence in the black hole interior*, *Phys. Rev.* **D93** (2016) 084049 [1503.08825].

[17] J. M. Maldacena, *Eternal black holes in anti-de Sitter*, *JHEP* **0304** (2003) 021 [hep-th/0106112].

[18] L. Susskind and Y. Zhao, *Teleportation Through the Wormhole*, 1707.04354.

[19] I. Heemskerk, D. Marolf, J. Polchinski and J. Sully, *Bulk and Transhorizon Measurements in AdS/CFT*, *JHEP* **1210** (2012) 165 [1201.3664].

[20] S. H. Shenker and D. Stanford, *Multiple Shocks*, *JHEP* **12** (2014) 046 [1312.3296].

[21] S. H. Shenker and D. Stanford, *Black holes and the butterfly effect*, *Journal of High Energy Physics* **3** (2014) 67 [1306.0622].

[22] M. Kac, *On the notion of recurrence in discrete stochastic processes*, *Bulletin of the American Mathematical Society* **53** (1947) 1002.

[23] S. Lloyd, *Ultimate physical limits to computation*, *Nature* **406** (2000) 1047 [9908043].

[24] I. Kourkoulou and J. Maldacena, *Pure states in the SYK model and nearly-AdS_2 gravity*, 1707.02325.

[25] K. Papadodimas, *A class of non-equilibrium states and the black hole interior*, 1708.06328.

[26] S. D. Mathur, *The Information paradox: A Pedagogical introduction*, *Class.Quant.Grav.* **26** (2009) 224001 [0909.1038].

[27] R. Bousso, *Firewalls From Double Purity*, 1308.2665.

[28] S. H. Shenker and D. Stanford, *Black holes and the butterfly effect*, 1306.0622.

[29] J. Maldacena, S. H. Shenker and D. Stanford, *A bound on chaos*, *JHEP* **08** (2016) 106 [1503.01409].

[30] P. Hayden and J. Preskill, *Black holes as mirrors: Quantum information in random subsystems*, *JHEP* **0709** (2007) 120 [0708.4025].

[31] A. Almheiri, A. Mousatov and M. Shyani, *Escaping the Interiors of Pure Boundary-State Black Holes*, 1803.04434.

[32] A. Almheiri, *Holographic Quantum Error Correction and the Projected Black Hole Interior*, 1810.02055. ;R. Jefferson, *Comments on black hole interiors and modular inclusions*, 1811.08900. ;S. Cooper, M. Rozali, B. Swingle, M. Van Raamsdonk, C. Waddell and D. Wakeham, *Black Hole Microstate Cosmology*, 1810.10601.

[33] R. Brustein and Y. Zigdon, *Revealing the interior of black holes out of equilibrium in the Sachdev-Ye-Kitaev model*, *Phys. Rev.* **D98** (2018) 066013 [1804.09017].

[34] G. 't Hooft, *Dimensional reduction in quantum gravity*, *Conf. Proc.* **C930308** (1993) 284 [gr-qc/9310026].

[35] L. Susskind, *Three Lectures on Complexity and Black Holes*, 2018, 1810.11563.

[36] M. Srednicki, *The approach to thermal equilibrium in quantized chaotic systems*, *Journal of Physics A: Mathematical and General* **32** (1999) 1163.

[37] E. Verlinde and H. Verlinde, *Black Hole Entanglement and Quantum Error Correction*, *JHEP* **1310** (2013) 107 [1211.6913]. ;E. Verlinde and H. Verlinde, *Black Hole Information as Topological Qubits*, 1306.0516. ;E. Verlinde and H. Verlinde, *Behind the Horizon in AdS/CFT*, 1311.1137.

[38] R. Haag, *Local quantum physics: Fields, particles, algebras, 2nd ed.* Springer, 1992.

[39] J. M. Bardeen, B. Carter and S. W. Hawking, *The Four laws of black hole mechanics*, *Commun. Math. Phys.* **31** (1973) 161. ;R. M. Wald, *Black hole entropy is the Noether charge*, *Phys. Rev.* **D48** (1993) R3427 [gr-qc/9307038]. ;V. Iyer and R. M. Wald, *Some properties of Noether charge and a proposal for dynamical black hole entropy*, *Phys. Rev.* **D50** (1994) 846 [gr-qc/9403028]. ;S. Hollands and R. M. Wald, *Stability of Black Holes and Black Branes*, *Commun. Math. Phys.* **321** (2013) 629 [1201.0463]. ;D. L. Jafferis, A. Lewkowycz, J. Maldacena and S. J. Suh, *Relative entropy equals bulk relative entropy*, *JHEP* **06** (2016) 004 [1512.06431].

[40] W. Donnelly and S. B. Giddings, *Diffeomorphism-invariant observables and their nonlocal algebra*, *Phys. Rev.* **D93** (2016) 024030 [1507.07921].

[41] A. Kitaev, "A Simple Model of Quantum Holography." http://online.kitp.ucsb.edu/online/entangled15/kitaev/.

[42] J. Polchinski and V. Rosenhaus, *The spectrum in the sachdev-ye-kitaev model*, *Journal of High Energy Physics* **2016** (2016) 1.

[43] J. Maldacena and D. Stanford, *Remarks on the sachdev-ye-kitaev model*, *Phys. Rev. D* **94** (2016) 106002.

[44] J. S. Cotler, G. Gur-Ari, M. Hanada, J. Polchinski, P. Saad, S. H. Shenker et al., *Black Holes and Random Matrices*, *JHEP* **05** (2017) 118 [1611.04650].

[45] J. Sonner and M. Vielma, *Eigenstate thermalization in the Sachdev-Ye-Kitaev model*, *JHEP* **11** (2017) 149 [1707.08013].

[46] I. Heemskerk, J. Penedones, J. Polchinski and J. Sully, *Holography from Conformal Field Theory*, *JHEP* **0910** (2009) 079 [0907.0151]. ;A. Fitzpatrick, E. Katz, D. Poland and D. Simmons-Duffin, *Effective Conformal Theory and the Flat-Space Limit of AdS*, *JHEP* **1107** (2011) 023 [1007.2412]. ;S. El-Showk and K. Papadodimas, *Emergent Spacetime and Holographic CFTs*, 1101.4163.

[47] C. Krishnan and K. V. P. Kumar, *Towards a Finite-N Hologram*, *JHEP* **10** (2017) 099 [1706.05364].

[48] D. J. Gross and V. Rosenhaus, *All point correlation functions in SYK*, *JHEP* **12** (2017) 148 [1710.08113].

[49] S. Lloyd, *Pure state quantum statistical mechanics and black holes*, *ArXiv e-prints* (2013) [1307.0378].

[50] V. Balasubramanian, B. Czech, V. E. Hubeny, K. Larjo, M. Rangamani et al., *Typicality versus thermality: An Analytic distinction*, *Gen.Rel.Grav.* **40** (2008) 1863 [hep-th/0701122].

[51] V. Balasubramanian, J. de Boer, V. Jejjala and J. Simon, *The Library of Babel: On the origin of gravitational thermodynamics*, *JHEP* **0512** (2005) 006 [hep-th/0508023].

[52] L. F. Alday, J. de Boer and I. Messamah, *The Gravitational description of coarse grained microstates*, *JHEP* **12** (2006) 063 [hep-th/0607222].

[53] V. Balasubramanian, J. de Boer, S. El-Showk and I. Messamah, *Black Holes as Effective Geometries*, *Class.Quant.Grav.* **25** (2008) 214004 [0811.0263].

[54] S. Raju and P. Shrivastava, *A Critique of the Fuzzball Program*, 1804.10616.

[55] A. L. Fitzpatrick, J. Kaplan and M. T. Walters, *Virasoro Conformal Blocks and Thermality from Classical Background Fields*, *JHEP* **11** (2015) 200 [1501.05315].

[56] C.-M. Chang, D. M. Ramirez and M. Rangamani, *Spinning constraints on chaotic large c CFTs*, 1812.05585.

[57] L. Foini and J. Kurchan, *Eigenstate thermalization hypothesis and out of time order correlators*, *Phys. Rev.* **E99** (2019) 042139 [1803.10658].

[58] G. J. Turiaci and H. Verlinde, *On cft and quantum chaos*, *Journal of High Energy Physics* **2016** (2016) 110.

[59] S. B. Giddings, *Nonviolent unitarization: basic postulates to soft quantum structure of black holes*, *JHEP* **12** (2017) 047 [1701.08765].

[60] P. Nayak, J. Sonner and M. Vielma, *Extended Eigenstate Thermalization and the role of FZZT branes in the Schwarzian theory*, 1907.10061.

[61] P. Nayak, J. Sonner and M. Vielma, *Eigenstate Thermalisation in the conformal Sachdev-Ye-Kitaev model: an analytic approach*, *JHEP* **10** (2019) 019 [1903.00478].

[62] S. Hawking, *Particle Creation by Black Holes*, *Commun.Math.Phys.* **43** (1975) 199.

[63] J. M. Maldacena, *The Large N limit of superconformal field theories and supergravity*, *Adv.Theor.Math.Phys.* **2** (1998) 231 [hep-th/9711200].

[64] O. Bratteli and D. W. Robinson, *Operator algebras and quantum statistical mechanics. 1. C* and W* algebras, symmetry groups, decomposition of states.* 1979.

[65] G. 't Hooft, *On the Quantum Structure of a Black Hole*, *Nucl.Phys.* **B256** (1985) 727.

[66] D. Harlow, *Aspects of the Papadodimas-Raju Proposal for the Black Hole Interior*, *JHEP* **11** (2014) 055 [1405.1995].

[67] S. W. Hawking, M. J. Perry and A. Strominger, *Soft Hair on Black Holes*, *Phys. Rev. Lett.* **116** (2016) 231301 [1601.00921].

[68] D. Kabat and G. Lifschytz, *Finite N and the failure of bulk locality: Black holes in AdS/CFT*, *JHEP* **09** (2014) 077 [1405.6394].

[69] J. Polchinski, L. Susskind and N. Toumbas, *Negative energy, superluminosity and holography*, *Phys. Rev.* **D60** (1999) 084006 [hep-th/9903228]. ;L. Susskind, *New Concepts for Old Black Holes*, 1311.3335.

[70] D. R. Brill and R. W. Lindquist, *Interaction energy in geometrostatics*, *Phys. Rev.* **131** (1963) 471.