

# UC Berkeley

## UC Berkeley Previously Published Works

### Title

Firewalls at exponentially late times

### Permalink

<https://escholarship.org/uc/item/6sp3m5w8>

### Journal

Journal of High Energy Physics, 2024(10)

### ISSN

1126-6708

### Authors

Blommaert, Andreas

Chen, Chang-Han

Nomura, Yasunori

### Publication Date

2024

### DOI

10.1007/jhep10(2024)131

### Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

# Firewalls at exponentially late times

Andreas Blommaert<sup>1</sup>, Chang-Han Chen<sup>2,3</sup> and Yasunori Nomura<sup>2,3,4,5</sup>

<sup>1</sup>SISSA and INFN, Via Bonomea 265, 34127 Trieste, Italy

<sup>2</sup>Berkeley Center for Theoretical Physics, Department of Physics,  
University of California, Berkeley, CA 94720, USA

<sup>3</sup>Theoretical Physics Group, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

<sup>4</sup>RIKEN iTHEMS, Wako, Saitama 351-0198, Japan

<sup>5</sup>Kavli IPMU (WPI), UTIAS, The University of Tokyo, Kashiwa, Chiba 277-8583, Japan

ablommae@sissa.it, changhanc@berkeley.edu, ynomura@berkeley.edu

## Abstract

We consider a version of the typical state firewall setup recently reintroduced by Stanford and Yang, who found that wormholes may create firewalls. We examine a late-time scaling limit in JT gravity in which one can resum the expansion in the number of wormholes, and we use this to study the exact distribution of interior slices at times exponential in the entropy. We consider a thermofield double with and without early perturbations on a boundary. These perturbations can appear on interior slices as dangerous high energy shockwaves. For exponentially late times, wormholes tend to teleport the particles created by perturbations and render the interior more dangerous. In states with many perturbations separated by large times, the probability of a safe interior is exponentially small, even though these would be safe without wormholes. With perturbation, even in the safest state we conceive, the odds of encountering a shock are fifty-fifty. One interpretation of the phenomenon is that wormholes can change time-ordered contours into effective out-of-time-ordered folds, making shockwaves appear in unexpected places.

# Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
1.1	Summary and structure . . . . .	3
<b>2</b>	<b>Thermofield double at exponentially late times</b>	<b>7</b>
2.1	Setup . . . . .	7
2.2	Exact answer from the gravitational path integral . . . . .	9
2.3	Semiclassical wavefunction and effective time . . . . .	11
2.4	Alternative derivation . . . . .	14
2.5	Gray holes . . . . .	15
<b>3</b>	<b>Simple perturbed thermofield double</b>	<b>17</b>
3.1	Logical overview . . . . .	18
3.2	Decomposing exact amplitude using effective times . . . . .	21
3.3	Semiclassical (scramblon) analysis of disk amplitude . . . . .	23
3.4	Conclusion . . . . .	29
3.5	Effective time-folds . . . . .	30
<b>4</b>	<b>Exponentially dangerous states</b>	<b>32</b>
4.1	Multiple shocks . . . . .	33
4.2	Firewall probabilities . . . . .	35
<b>5</b>	<b>Concluding remarks</b>	<b>37</b>
5.1	Room for improvement . . . . .	37
5.2	Perturbative firewall probability plateaus . . . . .	38
5.3	Lorentzian spacetimes . . . . .	39
<b>A</b>	<b>Avoided crossings</b>	<b>41</b>

# 1 Introduction

The AdS/CFT correspondence [1] has proven to be a very useful tool to understand the rules of quantum gravity. In particular, much recent progress on quantum gravity stems from attempting to find a bulk explanation for exotic physics at late times in the (often chaotic [2,3]) dual quantum system. Examples include the non-decaying (and highly oscillatory) behavior or late-time correlators of an eternal black hole [4] and unitarity of black hole evaporation embodied by the Page curve [5]. Both phenomena are explained on the AdS side by (spacetime) wormholes [6–14]. This is a category of questions which was easier to answer in the CFT than in AdS, which taught us the importance of wormholes.

However, other physically interesting questions about black holes in AdS have proven to be difficult to translate into simple questions in CFT. In particular, this includes the experiences of an observer falling into or residing in a black hole interior. Perhaps those questions are easier to tackle from the bulk point of view, relying on previous lessons on quantum gravity due to AdS/CFT. In this spirit, Stanford and Yang recently asked [15] if spacetime wormholes have an impact on the firewall question [16–18].

These authors considered the geometry dual [19] to the (time evolved) thermofield double (TFD), and found that a single wormhole has the potential to shorten (or lengthen) the Einstein–Rosen bridge (ERB), as will be shown in (1.3). This can even turn expanding time slices into contracting slices (the dual of the TFD at negative times). At times of order the Heisenberg time [20] (or inverse level spacing)<sup>1</sup>

$$T_H = 2\pi e^{S(E)}, \tag{1.1}$$

the correction due to the single wormhole amplitude to the total probability of finding an expanding or contracting slice, was found to compete with the leading disk contribution. This raises two immediate questions:

1. If the one-wormhole contribution competes with the no-wormhole one, then one should also consider contributions from any number of wormholes (and potentially non-perturbative corrections to that series). What is the final distribution of dual semiclassical geometries for  $T \sim T_H$ ?
2. Neither positive nor negative time slices of the TFD have firewalls in the setup that we will study (and with our definition of firewalls), which we will detail below. Potentially more dangerous states have matter perturbations on the boundary in the preparation of the state. Including wormholes, and given any initial state with matter perturbations, what are the odds of encountering a firewall at exponentially late times  $T \sim T_H$ ?

Like Stanford and Yang, we study these questions in 2d Jackiw–Teitelboim (JT) dilaton gravity [21–26]. As this theory does not include matter, there is not much that could endanger an observer crossing the horizon in the TFD. On the other hand, early matter perturbations in the preparation of the state

---

<sup>1</sup> This Heisenberg time is equal to the plateau time [3].

grow up to be high energy shockwaves in the bulk [2,27]. Collision with shocks with exponentially high energy *are* harmful. We include matter perturbations in our setup, and find that this can have major effects, deviating significantly from the result of Stanford and Yang [15]. The questions of black holes at exponentially late times were also analyzed by Susskind using complexity geometry [28].

We use the criterion that **bulk slices are dangerous if there is at least one strong shockwave in the slice**. By strong, we mean that the shock significantly affects the length of the dual slice. This is what we will mean by the (loaded) word “firewall,” nothing more.<sup>2</sup> We will see that wormholes, besides from changing the bare length of the slice [15], can also teleport the matter perturbations, resulting in shockwaves in unexpected places. We compute the resulting odds of having a firewall in the bulk slice.

On a technical level, we obtain all-genus results in a relatively straightforward way for two reasons. Firstly, unlike [15] we define our probability “operationally”; see (2.5). By this we mean they represent density matrices in which you compute expectation values of operators. After all, if we want to imagine that the slice is related to the experiences of an observer, then clearly some type of measurement on the slice must be imagined. This removes the mapping class group subtleties of [15]. Secondly, we will work in the “tau-scaling” limit [35–38]

$$T \rightarrow \infty, \quad T_H \rightarrow \infty, \quad T/T_H \text{ fixed}, \quad (1.2)$$

for all times  $T$  in the problem. In this regime, we can use exact results from random matrix theory [8,39] to account for summing over any number of wormholes in a simple manner. We will furthermore consider a regime where the Schwarzian can be treated semi-classically, as in [15]. In combination, this gives a geometric (semi-classical) interpretation of the full amplitudes.

While our setup evades the mapping class group subtleties of [15], it is not clear which setup most accurately models an infalling observer. Our setup has the advantage of being more closely related to a measurement that can be performed on boundaries, and allows on a technical level an incorporation of all-genus effects. The setup of [15] has the advantage that it has no formal divergences we encounter in section 2.5. To reach a more definite conclusion for the experience of an infalling observer, we would need a more realistic model of the observer.

## 1.1 Summary and structure

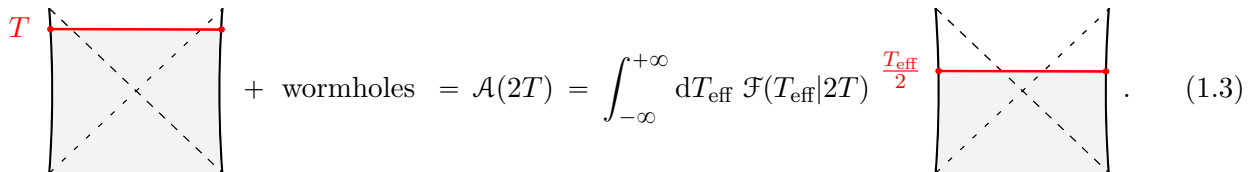
We now summarize the main results of our analysis.

In **section 2**, we study time evolution of the unperturbed TFD [15]. The classical dual geometry

---

<sup>2</sup> There are certainly other sensible definitions of firewalls and “experiments” to diagnose them, which would be interesting to investigate. We believe that our techniques could be modified without too much effort to different setups. An important issue which we do not explicitly address in this paper is the relation of our analyses to the property of the horizon of a collapse-formed single-sided black hole. The construction of [29–34] suggests that the effective two-sided black hole that emerges from such a black hole (after it stabilizes) is the one with  $T \ll T_H$  with essentially no perturbation, implying that an infalling observer would not encounter a firewall. We leave a closer investigation of this to the future.

(ignoring wormholes) is the time  $T$  slice of the two-sided black hole. Quantum mechanically (including wormholes), we instead find that there is a nonzero probability for the geometry to be the slice of the two-sided black hole with **effective age**  $T_{\text{eff}} \neq 2T$ . Schematically, the amplitude decomposes as<sup>3</sup>



$$+ \text{wormholes} = \mathcal{A}(2T) = \int_{-\infty}^{+\infty} dT_{\text{eff}} \mathcal{F}(T_{\text{eff}}|2T) \frac{T_{\text{eff}}}{2} \quad (1.3)$$

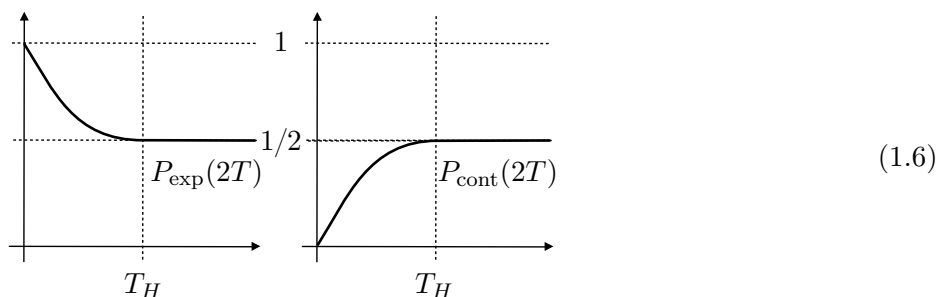
The “conversion factor”  $\mathcal{F}(T_{\text{eff}}|2T)$  captures the physical idea that wormholes can change the effective age of the ERB [7,15]. This amplitude has support also for negative effective times  $T_{\text{eff}} < 0$ , which slice through the white hole interior. One can ask [15] about the chances of finding a black hole (expanding slice) or white hole (contracting slice) as a function of  $T$ . This is the integrated conditional probability

$$P_{\text{exp}}(2T) = \int_0^{+\infty} dT_{\text{eff}} \mathcal{F}(T_{\text{eff}}|2T), \quad P_{\text{cont}}(2T) = \int_{-\infty}^0 dT_{\text{eff}} \mathcal{F}(T_{\text{eff}}|2T). \quad (1.4)$$

A version of this computation, including the correction from one wormhole, was carried out in [15] (see also [40]). We do non-perturbative calculation in the tau-scaling limit, which sums the contribution of any number of wormholes, and find that before the Heisenberg time  $2T < T_H = 2\pi e^{S(E)}$

$$P_{\text{exp}}(2T) = 1 - \frac{2T}{T_H} + \frac{1}{2} \frac{(2T)^2}{T_H^2}, \quad P_{\text{cont}}(2T) = \frac{2T}{T_H} - \frac{1}{2} \frac{(2T)^2}{T_H^2}. \quad (1.5)$$

Moreover, after a Heisenberg time  $2T > T_H$  one finds exactly the **gray hole** result anticipated in [15,41], where expanding and contracting slices are equally likely  $P_{\text{exp}}(2T) = P_{\text{cont}}(2T) = 1/2$ . Combining these piecewise functions leads to the following behavior:



Given the close mathematical relation with the famous plateau in the spectral form factor [3], we call the  $2T > T_H$  behavior in (1.6) the **firewall probability plateau**.

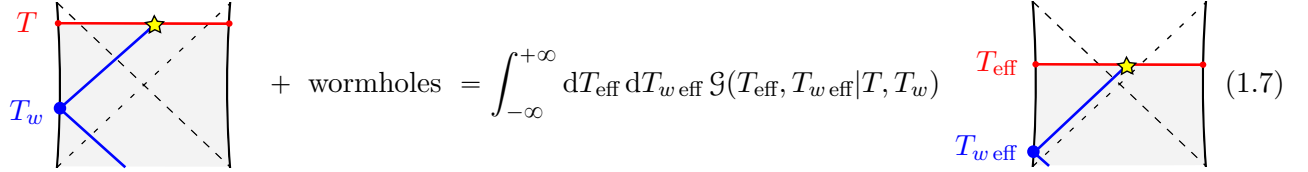
In **section 3**, we will study the TFD with one thermal perturbation at early times  $t = -T_w$  on the

---

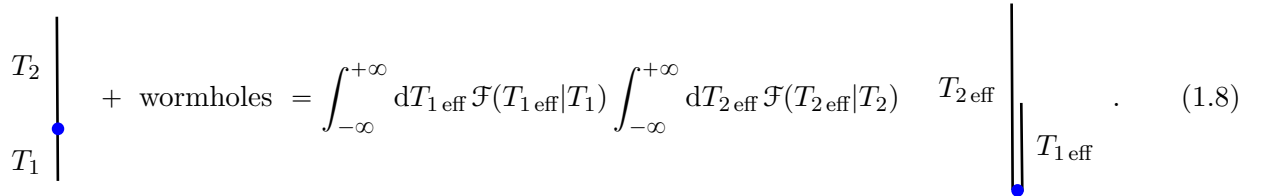
<sup>3</sup> We have only depicted the preparation of the ket, which will be glued onto an identical geometry preparing the bra at the red (interior) slice. The Euclidean preparation region is implicit. The factor of 2 between  $T$  and  $T_{\text{eff}}$  arises because  $T$  reflects two-sided time evolution while  $T_{\text{eff}}$  can be thought of as one-sided evolution. The latter will be more natural for later sections.

left boundary at Heisenberg time scales,  $T, T_w \sim T_H$ . In JT gravity, neither expanding nor contracting slices of the TFD are dangerous. Potential danger arises from high energy particles that cross the slices. In our setup, such particles may arise only from perturbations that were added in the preparation of the state.<sup>4</sup> We find that wormholes may teleport the perturbations far to the past or future, changing the **effective perturbation time**  $T_{w \text{ eff}} \neq T_w$ , and therefore potentially rendering naively safe encounters into dangerous ones (and vice versa).

We will show that the amplitude at time  $T$  decomposes (in our regime of approximation) as



with some conversion factor  $\mathcal{G}(T_{\text{eff}}, T_{w \text{ eff}} | T, T_w)$ . Introducing  $T_1 = T - T_w$  and  $T_2 = T + T_w$ , we find that the conversion factor factorizes to a product of conversion factors appearing in (1.3):  $\mathcal{G}(T_{\text{eff}}, T_{w \text{ eff}} | T, T_w) \sim \mathcal{F}(T_{1 \text{ eff}} | T_1) \mathcal{F}(T_{2 \text{ eff}} | T_2)$ . This has support for all signs of  $T_{1 \text{ eff}}, T_{2 \text{ eff}}$ , and therefore wormholes in this observable offer the possibility to essentially **turn time-ordered contours into effective out-of-time ordered folds**, and vice versa. Here we are referring to the time-folds that arise in the boundary dual when preparing the state at the late time slice of interest, starting from the  $t = 0$  TFD.<sup>5</sup> Adhering to the notation of [45], we can indeed write (1.7) schematically as



We use these time-folds as shorthand notation for classical slices being probed in the dual geometry.

We consider the interior slice to be **dangerous** if and only if there is a strong shockwave, meaning that a shock severely backreacts on the geometry of the effective slice.<sup>6</sup> This happens **if there is at least one switchback in the effective time-fold** preparing the state [45]. At very late times, all signs of  $T_{i \text{ eff}}$  are equally likely. The options of  $(T_{1 \text{ eff}}, T_{2 \text{ eff}})$  being  $(++)$  and  $(--)$  have no switchbacks. The options  $(+-)$  and  $(-+)$  have a switchback and are dangerous. Therefore, **the odds of encountering**

<sup>4</sup> We will consider JT gravity with probe matter, not a (dynamical) matter QFT coupled to JT gravity. It would be very interesting to study also the latter. Unfortunately, however, studying dynamical matter on wormholes in JT is notoriously challenging [8, 42] (but see [43, 44]). We will not attempt this here, but comment more on this in the discussion section 5.

<sup>5</sup> See equation (4.22) and Figure 6 in [45] or our equation (3.68).

<sup>6</sup> Technically, in our JT gravity setup the discriminator is whether or not the length of the slice significantly changes due to shocks.

a firewall at very late times are even in this setup

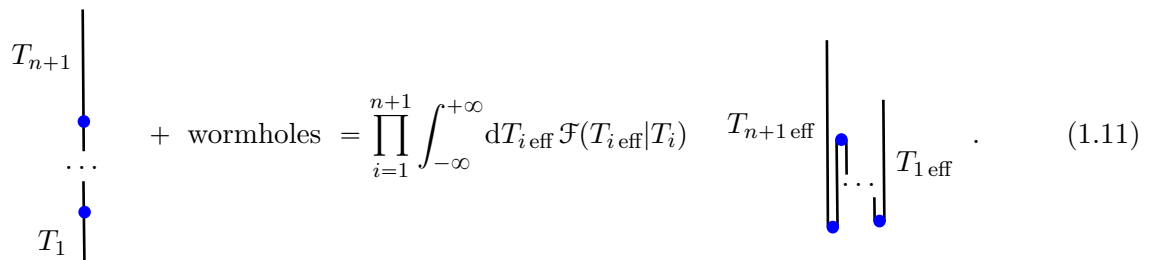
$$\boxed{P_{\text{safe}}(T) = P_{\text{danger}}(T) = \frac{1}{2}}, \quad T > T_H + T_w. \quad (1.9)$$

In **appendix A**, we show that another effect of wormholes, namely the possibility that the perturbation is carried away by a wormhole (thus avoiding a collision in the interior), vanishes in the tau-scaling limit.

In **section 4**, we consider more general states obtained by perturbing the TFD on the left boundary at multiple times  $t = -T_{w_i}$  ( $i = 1, \dots, n$ ), where  $T_{w_1} < \dots < T_{w_n}$ . We will consider several different configurations. The worst case scenario is one where  $n$  perturbations are each separated by more than a Heisenberg time, in which case **dangerous interiors are exponentially likely** in  $n$

$$\boxed{P_{\text{safe}}(T) = \frac{1}{2^n}} \xrightarrow{n \gg 1} 0, \quad T > T_{w_n} + T_H. \quad (1.10)$$

This is simple to see using the generalization of (1.8). The effective time-fold consists of  $n + 1$  segments



$$\begin{array}{c} T_{n+1} \\ | \\ \bullet \\ \vdots \\ | \\ \bullet \\ | \\ T_1 \end{array} + \text{wormholes} = \prod_{i=1}^{n+1} \int_{-\infty}^{+\infty} dT_{i,\text{eff}} \mathcal{F}(T_{i,\text{eff}}|T_i) \begin{array}{c} T_{n+1,\text{eff}} \\ | \\ \bullet \\ \vdots \\ | \\ \bullet \\ | \\ T_{1,\text{eff}} \end{array}. \quad (1.11)$$

For sufficiently late times, the sign of the times  $T_{i,\text{eff}}$  of each segment is random. The only safe scenarios are when there are no switchbacks, so all signs of  $T_{i,\text{eff}}$  must match. This is exponentially unlikely in  $n$ . We consider other states, but at late times none of them are safer than the single-particle setup (1.9), which is realized if the  $n$  perturbations are separated by times much less than the scrambling time  $T_S$ .

We stress again that what we are computing here is the probability that there is at least one dangerous shock on the interior slice in pure JT gravity. As shown in Figure 4 in [27], on a multiple shockwave geometry in higher dimensions, only the outermost shockwave determines the experience of an infalling observer. In this case, we would conclude that even with multiple shock waves, the probability of firewall at  $T \rightarrow \infty$  would still be  $P = 1/2$ .<sup>7</sup> In pure JT gravity, however, there is no spacelike singularity, so the probability of an infalling observer not being hit by a shock is (1.10).

Finally, in **section 5** we reemphasize that these results are obtained by summing a perturbatively convergent series (in genus) of wormhole amplitudes, following [37,38,46], and identify the corresponding Lorentzian wormhole geometries [47,48]. These explicitly realize the notion of effective times. We will also identify several shortcomings of our work, making concrete proposals for how to improve on it.

<sup>7</sup> We thank Douglas Stanford for pointing this out.



## Relation to other work

Similar questions have been independently investigated at the same time via related techniques by Iliesiu, Levine, Lin, Maxfield and Mezei [49], with whom we have coordinated submissions.

## 2 Thermofield double at exponentially late times

In this section, we study the effects of wormholes on the time evolution of the unperturbed TFD state in JT gravity. In particular, we will ask about the distribution of the length and bulk spatial slices as a function of boundary time  $T$ . From here on, we will always picture purely Euclidean spacetimes, with the appropriate Lorentzian continuation implied by the boundary conditions.

### 2.1 Setup

Suppose one prepares the TFD at  $t = 0$ , and time evolve this state on both sides during boundary time  $T$ . We are interested in the bulk interpretation of the resulting state. We imagine this is diagnosed by performing simple two-sided measurements, such as computing two-sided two-point functions.<sup>8</sup> In JT gravity those are computed by summing over wormholes [7, 9, 10, 12]<sup>9</sup>

$$G_{\Delta \text{ nonpert}}(2T) = \int_{\beta_2}^{\beta_1} \int_{\beta_1}^{\beta_2} \mathcal{O}_{\Delta} \mathcal{O}_{\Delta} + \int_{\beta_2}^{\beta_1} \int_{\beta_1}^{\beta_2} \mathcal{O}_{\Delta} \mathcal{O}_{\Delta} + \dots \quad (2.1)$$

with analytic continuation of the boundary conditions to Lorentzian times

$$\beta_1 = \frac{\beta}{2} + 2iT, \quad \beta_2 = \frac{\beta}{2} - 2iT. \quad (2.2)$$

Here, as announced, we have depicted the Euclidean  $\text{AdS}_2$  wormhole geometries. These path integrals can be computed exactly and the full amplitude decomposes as [7, 9, 10, 12]

$$G_{\Delta \text{ nonpert}}(2T) = \int_{-\infty}^{+\infty} d\ell \tilde{\mathcal{F}}(\ell|2T) G_{\Delta \text{ class}}(\ell), \quad (2.3)$$

where for boundary operators of conformal weight  $\Delta$  (dual to a bulk particle of mass  $\Delta$ ) [7, 50, 51]

$$G_{\Delta \text{ class}}(\ell) = e^{-\Delta \ell}. \quad (2.4)$$

<sup>8</sup> We could more generally consider two-sided higher point functions (with operators inserted at the same boundary time) which in the tau-scaling limit are computed using the same distribution  $\tilde{\mathcal{F}}(\ell|2T)$  in (2.3). Off-diagonal contributions in the  $\ell$ -basis are then suppressed by powers of  $e^{S_0}$ , since without exponentially large times weighting the boundary segments between operator insertions, there is nothing canceling such suppression.

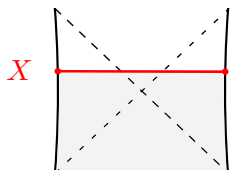
<sup>9</sup> The argument is  $2T$  because we are evolving both the left and right sides of the TFD by  $T$ .

Here  $\ell$  is the geodesic distance covered by the particle. In JT gravity this decomposition (2.3) is exact.

The quantity of our interest is a slightly modified version of the distribution  $\tilde{\mathcal{F}}(\ell|2T)$

$$G_{\Delta \text{ nonpert}}(2T) = \int_{-\infty}^{+\infty} dT_{\text{eff}} \mathcal{F}(T_{\text{eff}}|2T) G_{\Delta \text{ class}}(T_{\text{eff}}). \quad (2.5)$$

$T_{\text{eff}}$  is defined to be the length of the spatial slice in the classical two-sided black hole when *one of the boundaries* is evolved by  $T_{\text{eff}}$ , or equivalently both boundaries are evolved by  $T_{\text{eff}}/2$ . (This definition is more convenient in later sections.) To be precise, in JT gravity the time-evolving TFD spacetime has the metric and dilaton<sup>10</sup>



$$ds^2 = \frac{-dX^2 + d\sigma^2}{\sin(\sigma)^2}, \quad \Phi = E^{1/2} \frac{\cos(X)}{\sin(\sigma)}. \quad (2.6)$$

In these conformal coordinates, the location of the boundary changes as a function of boundary time  $t$  as (see for instance [25] for more details)

$$\tan(X/2) = \tanh(E^{1/2}t), \quad \sigma = \pi - \varepsilon \frac{dX}{dt}. \quad (2.7)$$

The geodesic between boundary points at  $t = T_{\text{eff}}/2$  is at constant  $X$  and has (renormalized) length

$$e^{\ell/2} = \frac{1}{E^{1/2}} \cosh(E^{1/2} T_{\text{eff}}). \quad (2.8)$$

The correct interpretation of (2.5) is that, after including wormhole corrections, the actual Lorentzian spacetime one is probing is a distribution of slices of the ordinary TFD with effective ages  $T_{\text{eff}}$  different from  $2T$ . One way to see this is to notice that  $G_{\Delta \text{ class}}(T_{\text{eff}})$  is the classical two-point function in such a spacetime. We will also see in section 2.3 that this indeed follows from a semiclassical analysis of the exact JT amplitudes, combined with a geometric interpretation of all integration parameters [51,53]. Finally, this interpretation also results because the purely Lorentzian analogues of the Euclidean wormhole geometries sketched in (2.1) are actually known [47], where the probed Lorentzian slices are indeed  $T_{\text{eff}}$  slices of the TFD. We will discuss those Lorentzian wormhole geometries in more detail in section 5.3.

Before we compute and analyze  $\mathcal{F}(T_{\text{eff}}|2T)$ , let us make two comments about our setup:

1. Our conversion factor  $\mathcal{F}(T_{\text{eff}}|2T)$ , even though very similar in spirit, is mathematically different

---

<sup>10</sup> Even though we use the terminology TFD we will almost exclusively study a microcanonical ensemble with fixed energy  $E$ . The Lorentzian spacetime for a fixed energy is also smooth [52] and classically  $E = \pi^2/\beta^2$ .

from the quantity  $\mathcal{F}_{\text{SY}}(T_{\text{eff}}|2T)$  computed by Stanford and Yang [15], who define

$$Z(2T) = \int_{-\infty}^{+\infty} d\ell \tilde{\mathcal{F}}_{\text{SY}}(\ell|2T) Z(\ell). \quad (2.9)$$

They decompose the partition function *without* operator insertions in the length basis. For famous mathematical reasons having to do with the mapping class group, their calculation is actually a more complicated problem in JT gravity than ours. This topic is very didactically documented, starting with [54–56]. How this plays together with a calculation of the two-point function in JT gravity was explained in [7, 9, 12], and reviewed nicely in [10]. We will not review it again here.

2. Our setup is at least as well motivated as that of Stanford and Yang. Indeed, without measurement there is not much physical relevance to amplitudes. The density matrix that is actually appearing in measurements is the amplitude we study in (2.5). We do not claim that the setup of [15] is not reasonable to study. As an additional a posteriori justification of our method, we note that our final answer (1.6) is physically sensible. We emphasize that we can not claim that infalling observers are modeled well by the simple two-sided measurements we consider here, even though this is the ultimate goal we must have in mind. See also section 5.

In the remainder of this section, we will compute and analyze  $\mathcal{F}(T_{\text{eff}}|2T)$ .

## 2.2 Exact answer from the gravitational path integral

We start with an exact expression for the amplitude of the two-point function in JT gravity [7, 9, 10, 12]

$$\begin{aligned} G_{\Delta \text{ nonpert}}(2T) &= \int_{-\infty}^{+\infty} d\ell \tilde{\mathcal{F}}(\ell|2T) e^{-\Delta\ell} \\ &= \frac{1}{Z(\beta)} \int_0^\infty dE_1 e^{-(\beta/2+2iT)E_1} \int_0^\infty dE_2 e^{-(\beta/2-2iT)E_2} \rho(E_1, E_2) e^{-S_0} \int_{-\infty}^{+\infty} d\ell \psi_{E_1}(\ell) \psi_{E_2}(\ell) e^{-\Delta\ell} \end{aligned} \quad (2.10)$$

with orthonormal wavefunctions

$$\psi_E(\ell) = 4K_{2iE^{1/2}}(e^{-\ell/2}), \quad \int_{-\infty}^{+\infty} d\ell \psi_{E_1}(\ell) \psi_{E_2}(\ell) = \frac{\delta(E_1 - E_2)}{\rho_0(E)}, \quad \rho_0(E) = \frac{\sinh(2\pi E^{1/2})}{4\pi^2}, \quad (2.11)$$

where  $S_0$  is the extremal entropy [8]. The integration measure  $\rho(E_1, E_2)$  or “spectral correlation” is

$$\rho(E_1, E_2) = \rho(E_1)\rho(E_2) + \rho(E_1, E_2)_{\text{conn}}, \quad (2.12)$$

with [8]

$$\rho(E) = e^{S_0} \rho_0(E) + \sum_{g=1}^{\infty} e^{(1-2g)S_0} \int_0^\infty db \rho_{\text{trumpet}}(E, b) V_{g,1}(b) + \text{non-perturbative}, \quad (2.13)$$

and

$$\begin{aligned} & \rho(E_1, E_2)_{\text{conn}} \\ &= \sum_{g=0}^{\infty} e^{-2gS_0} \int_0^{\infty} db_1 \rho_{\text{trumpet}}(E_1, b_1) \int_0^{\infty} db_2 \rho_{\text{trumpet}}(E_2, b_2) V_{g,2}(b_1, b_2) + \text{non-perturbative}. \end{aligned} \quad (2.14)$$

Here, the “trumpet” density of states is [8]

$$\rho_{\text{trumpet}}(E, b) = \frac{\cos(bE^{1/2})}{2\pi E^{1/2}}, \quad (2.15)$$

and  $V_{g,n}(b_1 \dots b_n)$  are Weil-Peterson volumes of the moduli spaces of Riemann surfaces [54–56].

The genus  $g$  expansions are completed non-perturbatively by a matrix integral answer [8]. However, if in addition to the large  $S_0$  limit, we also take the late-time limit, namely the tau-scaling limit, (1.2)

$$T \rightarrow \infty, \quad e^{S_0} \rightarrow \infty, \quad Te^{-S_0} \text{ fixed}, \quad (2.16)$$

then this story simplifies significantly. In fact, the genus expansion reproduces already (2.19) [15, 38]; *no* non-perturbative completion is needed! This will be the regime of interest in this paper.

To further elaborate, let us introduce the parametrization

$$E_1 = E + \frac{\omega}{2}, \quad E_2 = E - \frac{\omega}{2}, \quad (2.17)$$

so that the Boltzmann weights in (2.10) become  $e^{-\beta E} e^{-2i\omega T}$ . For  $T \sim e^{S_0} \rightarrow \infty$  the Fourier transform (the  $\omega$  integral) receives contributions from the least analytic components of  $\rho(E_1, E_2)$  as a function of  $\omega$ . They happen to arise from the tau-scaling limit in the energy domain

$$\omega \rightarrow 0, \quad e^{S_0} \rightarrow \infty, \quad \omega e^{S_0} \text{ fixed}. \quad (2.18)$$

This is precisely the regime [57] in which correlators in random matrix theory have a universal answer, featuring the so-called sine kernel [20, 39]

$$\rho(E) = e^{S_0} \rho_0(E), \quad \rho(E_1, E_2) = \rho(E)^2 + \delta(\omega) \rho(E) - \frac{\sin(\pi \rho(E) \omega)^2}{\pi \omega^2}, \quad (2.19)$$

where  $\rho(E) = e^{S(E)}$ . Importantly, the argument of the sine function contains a factor of  $e^{S_0}$  from the density of states. The generalization of (2.19) to multiple energy entries  $\rho(E_1 \dots E_n)$  is well known [39] and will be implicitly used in sections 3 and 4.<sup>11</sup>

We will work at fixed  $E$  instead of  $\beta$  (by doing an inverse Laplace transform) and consider large

---

<sup>11</sup> For detailed equations, see for instance (2.26) in [12] and section 3.2 in [58].

black holes  $E \gg 1$ , which suppresses Schwarzian (quantum) fluctuations. At this stage, we have

$$\tilde{\mathcal{F}}(\ell|2T) = \frac{e^{-S_0}}{\rho(E)} \int_{-\infty}^{+\infty} d\omega e^{-2i\omega T} \psi_{E+\omega/2}(\ell) \psi_{E-\omega/2}(\ell) \rho(E_1, E_2) \quad (2.20)$$

with the spectral correlation evaluated as (2.19). Below we will discuss the semiclassical approximation of the wavefunctions  $\psi_{E+\omega/2}(\ell) \psi_{E-\omega/2}(\ell)$ .<sup>12</sup>

### 2.3 Semiclassical wavefunction and effective time

We claim that the correct semiclassical approximation of the wavefunctions is<sup>13</sup>

$$\psi_{E+\omega/2}(\ell) \psi_{E-\omega/2}(\ell) = \Theta(\ell + \log(E)) \frac{1}{\rho_0(E)} \frac{1}{2\pi} \frac{1}{(E - e^{-\ell})^{1/2}} \cos\left(\frac{\omega}{E^{1/2}} \operatorname{arccosh}\left(2e^{\ell/2} E^{1/2}\right)\right). \quad (2.24)$$

Let us derive this. The integral representations of the Bessel functions gives

$$\begin{aligned} & \psi_{E+\omega/2}(\ell) \psi_{E-\omega/2}(\ell) \quad (2.25) \\ &= \int_{-\infty}^{+\infty} db_1 db_2 \exp\left(i(b_1 + b_2)E^{1/2} + i(b_1 - b_2)\frac{\omega}{4E^{1/2}} - 2e^{-\ell/2} \cosh\left(\frac{b_1 + b_2}{4}\right) \cosh\left(\frac{b_1 - b_2}{4}\right)\right). \end{aligned}$$

Introducing ‘‘angular’’ variables  $\alpha_i$  (the geometrical meaning of which we will discuss shortly)

$$2\pi + ib_i = \alpha_i \quad (2.26)$$

and furthermore relabeling these integration variables as

$$\alpha_1 = \pi + 2iE^{1/2}(T_{\text{eff}} + \Delta T_{\text{eff}}), \quad \alpha_2 = \pi + 2iE^{1/2}(-T_{\text{eff}} + \Delta T_{\text{eff}}), \quad (2.27)$$

---

<sup>12</sup> The result we will obtain from this is not qualitatively new. In section 2.2 of [15], the authors approximate the Bessel functions by cosines, resulting in contributions such as

$$\psi_{E+\omega/2}(\ell) \psi_{E-\omega/2}(\ell) \sim e^{i\omega \frac{\ell}{2\sqrt{E}}}, \quad (2.21)$$

Our approximation is more precise, but more importantly, it prepares us well for the more complicated setup of sections 3.

<sup>13</sup> An immediate check is the following computation in the tau-scaling limit

$$\begin{aligned} \int_{-\log(E)}^{+\infty} d\ell \frac{1}{\rho(E)} \frac{1}{2\pi} \frac{1}{(E - e^{-\ell})^{1/2}} \cos\left(\frac{\omega}{E^{1/2}} \operatorname{arccosh}\left(e^{\ell/2} E^{1/2}\right)\right) e^{-\Delta\ell} &= \int_{-\log(E)}^{+\infty} d\ell \frac{1}{\rho(E)} \frac{1}{2\pi} \frac{1}{(E - e^{-\ell})^{1/2}} e^{-\Delta\ell} \\ &= \frac{1}{2\pi} \frac{1}{\rho(E)} \frac{\Gamma(\Delta)^2}{\Gamma(2\Delta)} e^{-(\Delta-1/2)(-\log(4E))}, \quad (2.22) \end{aligned}$$

which matches the semiclassical approximation that one gets by directly taking  $E \gg 1$  and  $\omega \ll 1$  in the Gamma functions that arise [50, 51, 59] when we compute

$$\int_{-\infty}^{+\infty} d\ell \psi_{E_1}(\ell) \psi_{E_2}(\ell) e^{-\Delta\ell} \quad (2.23)$$

using the exact wavefunctions (2.11). In the second step we used the fact that the exponential suppression  $e^{-\Delta\ell}$  destroys contributions from lengths of order  $e^{S_0}$  such that the cosine evaluates to one.

the integrand becomes

$$\exp\left(-2\pi E^{1/2} + i\omega T_{\text{eff}} + 4iE\Delta T_{\text{eff}} - 2ie^{-\ell/2} \cosh\left(T_{\text{eff}}E^{1/2}\right) \sinh\left(\Delta T_{\text{eff}}E^{1/2}\right)\right). \quad (2.28)$$

Without wormhole corrections, represented by the second and third terms of  $\rho(E_1, E_2)$  in (2.19), one could at this point already do the  $\omega$  integral in (2.20) and find  $T_{\text{eff}} = 2T$ . This is the classical answer on the disk, where  $\alpha_i$  have the interpretation [51, 53] of angles on the Euclidean disk

$$\alpha_i = \frac{2\pi}{\beta} \beta_i. \quad (2.29)$$

In our notation,  $\beta_1 = \beta/2 + 2iT$  and  $\beta_2 = \beta/2 - 2iT$  with  $\pi/\beta = E^{1/2}$ . Then one indeed finds  $T_{\text{eff}} = 2T$  and  $\Delta T_{\text{eff}} = 0$ .

However, the corrections due to wormholes in the integration kernel, (2.19) in (2.20), mean that the  $\omega$  integral no longer localizes on  $T_{\text{eff}} = 2T$ . We find it more convenient to keep the  $\omega$  integral in (2.20) and instead look for saddles of the  $T_{\text{eff}}$  and  $\Delta T_{\text{eff}}$  (or  $b_1$  and  $b_2$ ) integrals. The  $T_{\text{eff}}$  and  $\Delta T_{\text{eff}}$  equations of motion are

$$\omega = 2e^{-\ell/2} E^{1/2} \sinh\left(T_{\text{eff}}E^{1/2}\right) \sinh\left(\Delta T_{\text{eff}}E^{1/2}\right), \quad e^{-\ell/2} = \frac{2 E^{1/2}}{\cosh\left(T_{\text{eff}}E^{1/2}\right) \cosh\left(\Delta T_{\text{eff}}E^{1/2}\right)}. \quad (2.30)$$

In the tau-scaling limit (2.18) where  $\omega \sim e^{-S_0} \rightarrow 0$ , this has the following solutions

$$\Delta T_{\text{eff}} = \frac{\omega}{4E^{3/2}} \rightarrow 0, \quad T_{\text{eff}} = \pm \frac{1}{E^{1/2}} \text{arccosh}\left(2e^{\ell/2} E^{1/2}\right), \quad (2.31)$$

resulting in the on-shell actions

$$\exp\left(-2\pi E^{1/2} \pm i \frac{\omega}{E^{1/2}} \text{arccosh}\left(2e^{\ell/2} E^{1/2}\right)\right). \quad (2.32)$$

After including the one-loop factors this results in (2.24). The Heaviside function arises from the fact that the saddle is only valid for real  $T_{\text{eff}}$ .

Since we are at fixed energy, we could do a change of coordinates from  $\ell$  to

$$T_\ell = \frac{1}{E^{1/2}} \text{arccosh}\left(2e^{\ell/2} E^{1/2}\right). \quad (2.33)$$

Then the wavefunction-squared would simplify tremendously<sup>14</sup>

$$d\ell \psi_{E+\omega/2}(\ell)\psi_{E-\omega/2}(\ell) = dT_\ell \Theta(T_\ell) \frac{1}{\rho_0(E)} \frac{1}{\pi} \cos(\omega T_\ell), \quad (2.35)$$

which we can check is still correctly normalized as in (2.11). We, however, see that it is more convenient to instead write the distribution in terms of  $T_{\text{eff}} = \pm T_\ell$ , which takes positive *and* negative values. In conclusion, we find that the conversion factor from  $T$  to  $T_{\text{eff}}$ , corresponding to that in (2.10) from  $T$  to  $\ell$ , is (semiclassically)

$$\mathcal{F}(T_{\text{eff}}|2T) = \frac{1}{2\pi\rho(E)^2} \int_{-\infty}^{+\infty} d\omega e^{i\omega(T_{\text{eff}}-2T)} \left( \rho(E)^2 + \delta(\omega)\rho(E) - \frac{\sin(\pi\rho(E)\omega)^2}{\pi\omega^2} \right). \quad (2.36)$$

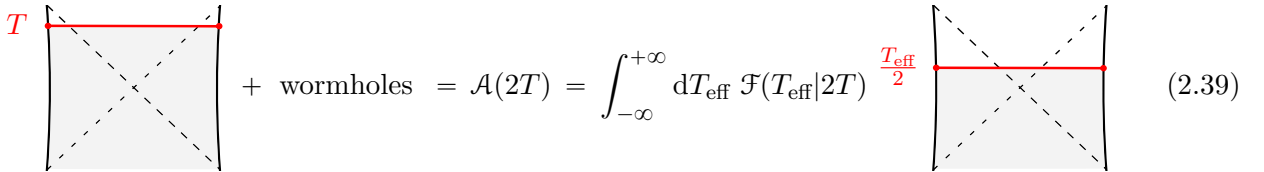
The disk contribution (the first term in the parentheses) is the classical answer. The bulk geometry is indeed the expected TFD with both sides evolved by  $T$ . In other words

$$\mathcal{F}(T_{\text{eff}}|2T) = \delta(T_{\text{eff}} - 2T) + \text{wormholes}. \quad (2.37)$$

We note that the contribution at fixed  $T_{\text{eff}}$  comes from the saddle

$$\alpha_1 = \pi + 2iE^{1/2} T_{\text{eff}}, \quad \alpha_2 = \pi - 2iE^{1/2} T_{\text{eff}}. \quad (2.38)$$

This is the saddle point of the disk path integral with Euclidean boundary times  $\beta_1 = \beta/2 + iT_{\text{eff}}$  and  $\beta_2 = \beta/2 - iT_{\text{eff}}$  on the segments between operator insertions. We will encounter a similar phenomenon in sections 3 and 4, where physics (the exact amplitude) essentially factorizes into on one hand wormhole physics, which changes the saddles of  $\alpha_i$  away from their classical disk answers, and on the other hand particle scattering at fixed  $\alpha_i$ . The latter reproduces a disk amplitude *as if* we would have put boundary conditions  $\beta_{i\text{eff}} = \alpha_i\beta/2\pi$  (with  $\alpha_i$  affected by wormholes and integrated over). In this case, this results in the pictorial representation of the amplitude



$$T \left[ \text{diagram} \right] + \text{wormholes} = \mathcal{A}(2T) = \int_{-\infty}^{+\infty} dT_{\text{eff}} \mathcal{F}(T_{\text{eff}}|2T) \frac{T_{\text{eff}}}{2} \left[ \text{diagram} \right] \quad (2.39)$$

with  $\mathcal{F}(T_{\text{eff}}|2T)$  given by (2.36). This is the expression quoted in (1.3) in the introduction.

We read this equation as showing us the Lorentzian way to interpret Euclidean wormhole calcula-

<sup>14</sup> As an intermediate check on our approximations here, note that at fixed  $\omega$  the expectation value of  $T_\ell$  is

$$\int_{-\infty}^{+\infty} dT_\ell \Theta(T_\ell) \frac{1}{\rho_0(E)} \frac{1}{\pi} \cos(\omega T_\ell) T_\ell = -\frac{1}{\pi} \frac{1}{\rho(E)} \text{fp} \left( \frac{1}{\omega^2} \right), \quad (2.34)$$

where fp is the Hadamard finite part. This matches exactly with equation (4.5) in [10].

tions. In other words, we interpret geometries with effective age  $T_{\text{eff}}$  as true Lorentzian slices that are being probed when we compute boundary two-point functions in JT gravity, and we want to know the distribution of such dual slices. Let us make several comments regarding this interpretation.

1. In the exact answer (2.10),  $\ell$  is the length of a geodesic in  $\text{AdS}_2$  to all orders in the genus expansion. A boundary observer measuring this correlator could reach two conclusions. They may take the fact that the answer does not match  $e^{-\Delta\ell(2T)}$  as evidence that any notion of classical spacetimes has failed. We believe that this would be a *wrong* conclusion, as the exact Euclidean calculation (in the tau-scaling limit) is purely geometric [38]. Or they could conclude that they are probing a wavefunction with support on different slices, each of which *does* make classical sense. All Lorentzian slices that can emerge in JT gravity are slices of the TFD, but at arbitrary times. This is (2.39). They find the wavefunction by decomposing the full wavefunction into basis functions  $e^{-\Delta\ell(T_{\text{eff}})}$  using (2.39), as we are doing here.
2. As already emphasized, we are not claiming that our discussion is directly applicable to describing the experiences of infalling observers. It is generally not well understood how to describe bulk observers in gravitational systems.<sup>15</sup> In particular, we do not really know that an infalling observer would experience the effective geometries on the right-hand sides of (1.3) and (1.7). Even though this seems plausible, it is an important open problem.
3. The reason to consider contributions to the integrals from  $T_{\text{eff}} \neq 2T$  is that the effect of wormholes introduces the sine kernel in (2.36), which has contributions with very large times like  $\sim e^{\pm i\omega T_H}$ . Such Fourier components generate contributions to  $\mathcal{F}(T_{\text{eff}}|2T)$  when  $T_{\text{eff}} - 2T = \pm T_H$ . For early times ( $T, T_{\text{eff}} \ll T_H$ ), the Fourier transform probes the coarse features of  $\rho(E_1, E_2)$ , and the delta function and sine kernel in (2.36) cancel each other out. This is the reason why wormholes are negligible for early time correlators, as they should because our everyday experiences clearly do not involve wormholes. But for very late times  $T \sim e^{S_0}$ , the wormhole corrections *can* compete, and dominate. For instance for two-point functions (2.10), the  $\rho(E)^2$  contribution will decay to zero in  $T \rightarrow \infty$  for any  $\Delta$  (as there are no infinitely sharp features in the integrand as function of  $\omega$ ). The delta and sine functions in (2.19) give rise to a non-decaying ramp-plateau contribution [3, 6, 7, 9, 15, 38] (which indeed vanishes for  $T \ll e^{S_0}$ ).

## 2.4 Alternative derivation

Here we provide a quick derivation for  $\mathcal{F}(T_{\text{eff}}|2T)$  (2.36). The effective-action analysis we just performed will be useful later in section 3.3, while the quick derivation here is more similar to the one in section 3.2.

---

<sup>15</sup> For some interesting recent progress, see [60, 61].



Starting with the exact, non-perturbative two-point function (2.10), one can write suggestively

$$G_{\Delta \text{ nonpert}}(2T) = \int_{-\infty}^{\infty} d\omega e^{2i\omega T} \frac{\rho(E_1, E_2)}{\rho(E_1)\rho(E_2)} G_{\Delta \text{ class}}(E_1, E_2). \quad (2.40)$$

Then we may simply resort to the convolution theorem

$$\int_{-\infty}^{+\infty} d\omega e^{2i\omega T} f(\omega) h(\omega) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} dT_{\text{eff}} \tilde{f}(2T - T_{\text{eff}}) \tilde{h}(T_{\text{eff}}), \quad (2.41)$$

where the tilded functions are individual Fourier transforms

$$\tilde{f}(2T - T_{\text{eff}}) = \int_{-\infty}^{+\infty} d\omega e^{i\omega(2T - T_{\text{eff}})} f(\omega), \quad \tilde{h}(2T_{\text{eff}}) = \int_{-\infty}^{+\infty} d\omega e^{2i\omega T_{\text{eff}}} h(\omega). \quad (2.42)$$

Applying this to our case with  $f(\omega) = \frac{\rho(E_1, E_2)}{\rho(E_1)\rho(E_2)}$  and  $h(\omega) = G_{\Delta \text{ class}}(E_1, E_2)$  reproduces (2.5) with

$$\mathcal{F}(T_{\text{eff}}|2T) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} d\omega e^{i\omega(T_{\text{eff}} - 2T)} \left( 1 + \frac{\delta(\omega)}{\rho(E)} - \frac{\sin(\pi\rho(E)\omega)^2}{\pi\omega^2\rho(E)^2} \right). \quad (2.43)$$

## 2.5 Gray holes

Given the simplified semiclassical expression (2.36) for  $\mathcal{F}(T_{\text{eff}}|2T)$  in the tau-scaling limit, it is simple to compute the all-genus probability of finding an expanding slice  $T_{\text{eff}} > 0$  or a contracting slice  $T_{\text{eff}} < 0$ , as function of boundary time  $2T$

$$P_{\text{exp}}(2T) = \int_0^{+\infty} dT_{\text{eff}} \mathcal{F}(T_{\text{eff}}|2T), \quad P_{\text{cont}}(2T) = \int_{-\infty}^0 dT_{\text{eff}} \mathcal{F}(T_{\text{eff}}|2T). \quad (2.44)$$

As an intermediate step, doing the Fourier transform in (2.36) leads to a shifted version of the ramp-and-plateau structure [3]

$$\boxed{\mathcal{F}(T_{\text{eff}}|2T) = \delta(T_{\text{eff}} - 2T) + \frac{1}{T_H^2} \min(|T_{\text{eff}} - 2T|, T_H)}, \quad T_H = 2\pi\rho(E), \quad (2.45)$$

which is constant after the Heisenberg [20] or plateau [3] time. This results for  $2T < T_H$  in the profile

$$P_{\text{exp}}(2T) = 1 - \frac{2T}{T_H} + \frac{1}{2} \frac{(2T)^2}{T_H^2} + \text{constant}, \quad P_{\text{cont}}(2T) = \frac{2T}{T_H} - \frac{1}{2} \frac{(2T)^2}{T_H^2} + \text{constant}, \quad (2.46)$$

and for  $2T > T_H$

$$P_{\text{exp}}(2T) = \frac{1}{2} + \text{constant}, \quad P_{\text{cont}}(2T) = \frac{1}{2} + \text{constant}, \quad (2.47)$$

where the infinite constant (on which we comment soon) is

$$\text{constant} = \frac{1}{T_H} \int_0^{\infty} dT_{\text{eff}} - \frac{1}{2}. \quad (2.48)$$

One noteworthy comment about this result (2.46) is that it agrees at genus one (up to a minus sign) with the results of Stanford and Yang [15], who found the quadratic piece  $(2T)^2/2T_H^2$ . The linear term is geometrically more mysterious, very much like certain terms in the Taylor series in  $2T$  of the interior length computed in [10]. Even though, as explained in section 2.1, our setup is mathematically different from [15] (due to our treatment of the mapping class group), it is comforting to find this agreement.

Regarding the infinite constant, we believe that the physically correct procedure is to subtract it. More precisely, the amplitude must be *renormalized*, by subtracting the values at  $T = 0$ . Indeed, the definition of our setup is that we start with the TFD at  $T = 0^+$ . Then we ask what happens to this geometry when one time evolves. To interpret  $P_{\text{exp,cont}}$  as probabilities, this means we should impose

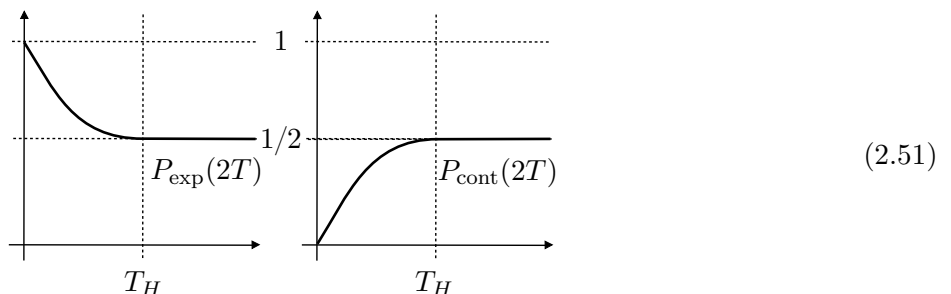
$$P_{\text{exp}}(2T = 0^+) = 1, \quad \text{and} \quad P_{\text{cont}}(2T = 0^+) = 0. \quad (2.49)$$

Subtracting this constant boils down to subtracting the ramp-plateau at  $2T = 0$  in  $\mathcal{F}(T_{\text{eff}}|2T)$  (2.45).<sup>16</sup> An identical renormalization was carried out when computing the interior length in [10]. Nevertheless, this remains a subtle point, one which we will come back to in the discussing section 5.3.

After this renormalization, one finds that our probabilities are correctly normalized

$$P_{\text{exp}}(2T) + P_{\text{cont}}(2T) = 1 \quad (2.50)$$

for all  $T$ . The resulting piecewise behavior as function of  $2T$  makes physical sense



In particular, expanding and contracting branches plateau at equal odds for post-Heisenberg times

$$\boxed{P_{\text{exp}}(2T) = P_{\text{cont}}(2T) = \frac{1}{2}}, \quad 2T > T_H. \quad (2.52)$$

This realizes the gray hole scenario anticipated in [15, 41]. For  $2T \ll T_H$  wormholes should be irrelevant for all practical purposes, and indeed the geometry is purely expanding, with the transition amplitude

---

<sup>16</sup> If we use different infrared cutoffs  $\Lambda_1$  and  $-\Lambda_2$  for the upper and lower edges of the integrals in (2.44), where  $\Lambda_{1,2}$  are (much) larger than any other scales in the problem, then the constants for  $P_{\text{exp}}$  and  $P_{\text{cont}}$  in (2.46) and (2.47) (which are now regularized) would be different. Even in this case, the condition (2.49) determines the necessary subtractions, so the final result (2.51) is not affected.

only having support on the disk contribution (which gives  $T_{\text{eff}} = 2T$  (2.37))

$$P_{\text{exp}}(2T) = 1, \quad P_{\text{cont}}(2T) = 0, \quad 2T \ll T_H. \quad (2.53)$$

We emphasize that in (2.46) the corrections due to wormholes occur at *leading* order. Even though the ramp-plateau in  $\mathcal{F}(T_{\text{eff}}|2T)$  is suppressed by  $1/T_H \sim e^{-S_0}$  in (2.45), we integrate  $T_{\text{eff}}$  over a range  $\sim T_H$  (the length of the ramp) when we compute the probabilities, creating competition at leading order!

As discussed in the introduction, in pure JT gravity, neither expanding nor contracting branch of the pure TFD is dangerous. In the remainder of this work we consider more generic states with early perturbations, which *could* be dangerous.<sup>17</sup>

### 3 Simple perturbed thermofield double

In this section, we study the simplest state that may have a dangerous interior, the TFD with a thermal perturbation at  $t = -T_w$  on the left boundary. We choose to focus again on exponentially large times  $T_w \sim T_H$ . The naive dual bulk slice at  $T \sim T_H$  (red) is



The early perturbation creates a shockwave [2], which depending on the values of  $T$  and  $T_w$  can be highly disruptive in the slice at time  $T$ . In particular, in JT gravity the geometry of the slice is characterized by its (renormalized) length  $\ell$  (2.8). For exponentially late times, to leading order in  $e^{S_0}$ , the length of an unperturbed slice is

$$\ell_{\text{bare}} = 2E^{1/2}|2T|. \quad (3.2)$$

In the presence of a shockwave, this may get modified. We *define* a strong shockwave to be one which affects this leading order behavior. We *define* slices with a strong shocks to be dangerous.

At disk level (i.e. ignoring wormholes), the presence of the shock modifies (3.2) to [45]<sup>18</sup>

$$\ell = 2E^{1/2}(|T_1| + |T_2|), \quad T_1 = T - T_w, \quad T_2 = T + T_w. \quad (3.3)$$

<sup>17</sup> An even more honest approach would be to treat matter as dynamical quantum fields, thereby including virtual processes such as the vacuum fluctuations. See section 5.1 for more discussion on this improvement.

<sup>18</sup> We will show how this is reproduced from a detailed analysis of the exact JT disk amplitude in section 3.3.

Comparing with (3.2), this leads to the naive conclusion [45] that the slice obeys<sup>19</sup>

$$\begin{aligned} |T_w| > |T| & \text{ dangerous,} \\ |T_w| < |T| & \text{ safe.} \end{aligned} \quad (3.4)$$

The contribution from wormholes is expected to change this naive conclusion. In this section, we want to answer the following question: including the effects of wormholes, what is the probability that the dual bulk slice is dangerous (meaning it contains a strong shockwave)?

### 3.1 Logical overview

In similar spirit to what we discussed in section 2.1, we start our analysis with the boundary four-point function in JT gravity, computed by summing over wormholes as [12]

$$\begin{aligned} G_{\Delta \Delta_w \text{ nonpert}}(T_1, T_2) = & \text{ (Diagram 1) } + \text{ (Diagram 2) } \\ & + \text{ (Diagram 3) } + \text{ (Diagram 4) } + \dots \end{aligned} \quad (3.5)$$

with analytic continuation of the boundary conditions to<sup>20</sup>

$$\beta_1 = \frac{\beta}{4} + \frac{\beta}{2\pi}\alpha + iT_1, \quad \beta_2 = \frac{\beta}{4} - \frac{\beta}{2\pi}\alpha + iT_2, \quad \beta_3 = \frac{\beta}{4} - \frac{\beta}{2\pi}\alpha - iT_2, \quad \beta_4 = \frac{\beta}{4} + \frac{\beta}{2\pi}\alpha - iT_1. \quad (3.6)$$

In section 3.2, we show that the out-of-time-order correlator (OTOC) on this contour (3.6) exactly decomposes, in the tau-scaling limit, as

$$G_{\Delta \Delta_w \text{ nonpert}}(T_1, T_2) = \int_{-\infty}^{+\infty} dT_{1 \text{ eff}} \mathcal{F}(T_{1 \text{ eff}}|T_1) \int_{-\infty}^{+\infty} dT_{2 \text{ eff}} \mathcal{F}(T_{2 \text{ eff}}|T_2) G_{\Delta \Delta_w \text{ disk}}(T_{1 \text{ eff}}, T_{2 \text{ eff}}). \quad (3.7)$$

One key step is to show that the dominant wormholes are empty wormholes that bridge over the (red) observer; precisely the ones pictured in (3.5). Other possibilities are wormholes which bridge over the

<sup>19</sup> The separation between  $|T_w|$  and  $|T|$  is assumed exponentially large in  $S_0$ .

<sup>20</sup> The genuinely Lorentzian setup (3.1) has  $\alpha = \pi/2$ , such that  $\text{Re } \beta_1 = \text{Re } \beta_4 = \beta/2$  and  $\text{Re } \beta_2 = \text{Re } \beta_3 = 0$ . We allow Euclidean separation between all operators, to avoid infinite energies. The precise value of  $\alpha$  will not affect our conclusions.

(blue) perturbation; and non-empty wormholes which make the perturbation bypasses the interior slice:

$$G_{\Delta\Delta_w \text{ nonpert}}(T_1, T_2) \supset \text{Diagram 1} + \text{Diagram 2} + \dots \quad (3.8)$$

We find that these are *negligible* in the tau-scaling limit (1.2). To be clear, every wormhole is exponentially suppressed in entropy by a factor of  $e^{-2S_0}$ , but some are amplified at exponentially late times due to the integrals over moduli  $T_{i \text{ eff}}$ . The second term in (3.8) is the “safest” geometry (the perturbation bypasses the observer). This term, however, does not have such moduli, and hence no enhancement in time. The kernels appearing in (3.7) notably have precisely the same universal form as (2.45)

$$\mathcal{F}(T_{i \text{ eff}}|T_i) = \delta(T_{i \text{ eff}} - T_i) + \frac{1}{T_H^2} \min(|T_{i \text{ eff}} - T_i|, T_H), \quad T_H = 2\pi\rho(E). \quad (3.9)$$

An important point is that, unlike the previous case where  $G_{\Delta \text{ class}} = e^{-\Delta\ell}$ , the OTOC, even at the disk level, does *not* always factorize

$$G_{\Delta\Delta_w \text{ disk}}(T_{1 \text{ eff}}, T_{2 \text{ eff}}) = \text{Diagram} \neq e^{-\Delta\ell} e^{-\Delta_w \ell_w}, \quad (3.10)$$

where  $\ell_w$  is the renormalized geodesic length between the two  $\mathcal{O}_{\Delta_w}$  insertions. The correct answer is given by the Schwarzian (or disk) OTOC four-point function (for fixed energy), computed for instance in [50,62] (and semiclassically in [25]). Even though this OTOC has been well studied in the literature, we conduct a more careful semiclassical (“scramblon” [53]) analyses in **section 3.3** to decompose it as

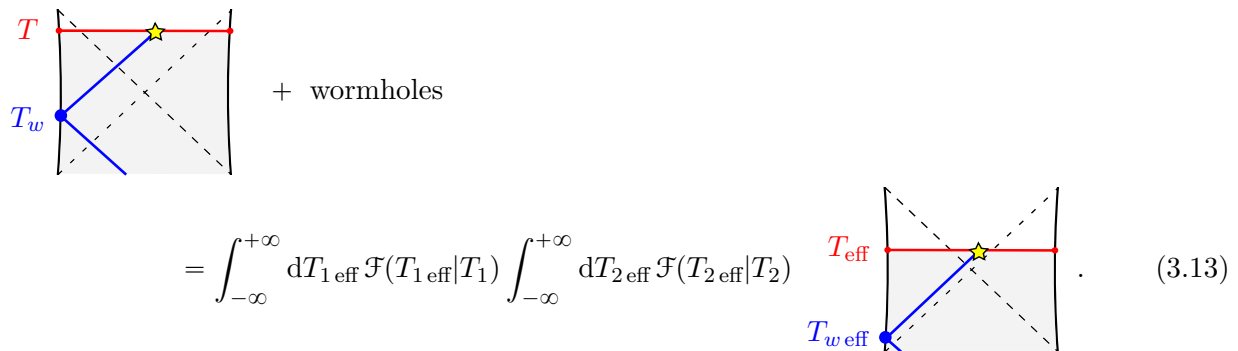
$$G_{\Delta\Delta_w \text{ disk}}(T_{1 \text{ eff}}, T_{2 \text{ eff}}) = \int_{-\infty}^{+\infty} d\ell \mathcal{A}_{\Delta_w \text{ disk}}(T_{1 \text{ eff}}, T_{2 \text{ eff}}, \ell) e^{-\Delta\ell}, \quad (3.11)$$

where  $\ell$  is the length of the interior slice. As in (2.3) we identify the kernel in (3.11) as the amplitude that prepares the slice with length  $\ell$  (and one perturbation). This results in the relation<sup>21</sup>

$$\mathcal{A}_{\Delta_w \text{ nonpert}}(T_1, T_2, \ell) = \int_{-\infty}^{+\infty} dT_{1 \text{ eff}} \mathcal{F}(T_{1 \text{ eff}}|T_1) \int_{-\infty}^{+\infty} dT_{2 \text{ eff}} \mathcal{F}(T_{2 \text{ eff}}|T_2) \mathcal{A}_{\Delta_w \text{ disk}}(T_{1 \text{ eff}}, T_{2 \text{ eff}}, \ell). \quad (3.12)$$

<sup>21</sup> We normalize this amplitude as follow. Imagine that there are no insertion  $\mathcal{O}_{\Delta}$  at the time slice  $T$ . Then the amplitude is equal to the two-point function of the perturbation  $\langle \mathcal{O}_{\Delta_w}(\beta_2 + \beta_3) \mathcal{O}_{\Delta_w}(0) \rangle$ . We divide the amplitude by this two-point function so that it is normalized to one. In (3.45), we will confirm that this normalization is indeed time-independent.

This leads to equation (1.7) in the introduction, with  $T_1 = T - T_w$  and  $T_2 = T + T_w$



$$= \int_{-\infty}^{+\infty} dT_{1\text{ eff}} \mathcal{F}(T_{1\text{ eff}}|T_1) \int_{-\infty}^{+\infty} dT_{2\text{ eff}} \mathcal{F}(T_{2\text{ eff}}|T_2) \cdot \quad (3.13)$$

The final step is to figure out what “safe” and “dangerous” means in the effective, classical geometry. In **section 3.3**, starting from the exact disk OTOC in JT gravity, we show that the normalized version of  $\mathcal{A}_{\Delta_w \text{ disk}}(T_{1\text{ eff}}, T_{2\text{ eff}}, \ell)$  simply reproduces the classical expectation of [45], with  $T_{1\text{ eff}} = T_{\text{eff}} - T_{w\text{ eff}}$  and  $T_{2\text{ eff}} = T_{\text{eff}} + T_{w\text{ eff}}$

$$\mathcal{A}_{\Delta_w \text{ disk}}(T_{1\text{ eff}}, T_{2\text{ eff}}, \ell) = \delta(\ell - 2E^{1/2}(|T_{1\text{ eff}}| + |T_{2\text{ eff}}|)). \quad (3.14)$$

Following (3.4), we conclude from (3.14) that the effective classical bulk slice being probed is dangerous respectively safe if

$ T_{w\text{ eff}}  >  T_{\text{eff}} $	dangerous
$ T_{w\text{ eff}}  <  T_{\text{eff}} $	safe

$$\cdot \quad (3.15)$$

Putting everything together, in **section 3.4** we compute the probability that the dual bulk slice is safe/dangerous, by integrating (3.12) over the regions (3.15). The  $\ell$  integral is trivial, due to the delta function (3.14), resulting in

$$P_{\text{safe}}(T_1, T_2) = \int_{\text{safe}} dT_{1\text{ eff}} dT_{2\text{ eff}} \mathcal{F}(T_{1\text{ eff}}|T_1) \mathcal{F}(T_{2\text{ eff}}|T_2). \quad (3.16)$$

As an aside, we note that there might be other sensible criteria besides (3.4) to distinguish whether or not the slice is dangerous. Our techniques allow for extending any replacement of the classical criterion (3.4) to the corresponding non-perturbative criterion (and resulting safe/danger probabilities) in a straightforward manner.

In **section 3.5**, we rephrase (3.15) using effective time-folds, as advertised in (1.8). This results in

effective out-of-time-ordered	dangerous
effective time-ordered	safe

$$\cdot, \quad (3.17)$$

which, hopefully, is more intuitive. We emphasize that it is the *effective* time-folds that are in one-to-one correspondence with the *classical spacetime*, not the boundary time-folds (which serve as the boundary condition for the exact calculation).

Sections 3.2 and 3.3 are technical. Readers interested only in physical results can skip to section 3.4.

### 3.2 Decomposing exact amplitude using effective times

Except for contributions such as the last diagram in (3.8) (which we address in section 3.4), the exact four-point function in JT gravity (3.5) decomposes as [12, 51]

$$G_{\Delta \Delta_w \text{ nonpert}}(T_1, T_2) \tag{3.18}$$

$$= \prod_{i=1}^4 \int_0^\infty \frac{dz_i}{z_i} \int_0^\infty dE_i \psi_{E_i}(z_i) e^{-\beta_i E_i} \rho(E_1, E_2, E_3, E_4) \int_0^\infty \frac{dz}{z} I_3(z_1, z_2, z) I_3(z_3, z_4, z) e^{-\Delta_w \ell_w} e^{-\Delta \ell}.$$

Here, we introduced the notation

$$z_i = e^{-\ell_i/2}, \quad z = e^{-\ell/2}, \tag{3.19}$$

where  $\ell_i$  is the renormalized geodesic length between the  $i$ -th and  $(i+1)$ -th insertions on the boundary. The four-level spectral density  $\rho(E_1, E_2, E_3, E_4)$  arises from the sum over wormholes and should be computed in the tau-scaling limit as the generalization of (2.19). Furthermore,

$$I_3(z_1, z_2, z_3) = \int dE \rho_0(E) \prod_{i=1}^3 \psi_E(z_i) = \exp\left(-\frac{1}{2} \frac{z_1 z_2}{z_3} - \frac{1}{2} \frac{z_2 z_3}{z_1} - \frac{1}{2} \frac{z_3 z_1}{z_2}\right). \tag{3.20}$$

Finally, the length  $\ell_w$  of the geodesic of the perturbation follows from some hyperbolic geometry

$$e^{\ell_w/2} = \frac{1}{z_w} = \frac{z}{z_1 z_3} + \frac{z}{z_2 z_4}. \tag{3.21}$$

The derivation is somewhat elaborate, but well documented. We here only summarize the main steps leading to (3.18) in words, and refer the interested reader to the relevant literature for details.

1. By using the correspondence between JT gravity and 2d topological Yang-Mills theory (also called BF theory), and thinking carefully about the mapping class group, one derives a version of (3.18) that, instead of all the  $z_i$  and  $z$  dependence, has a kernel which (aside from the four-level spectral density) involves a 6j symbol of  $\text{SL}(2, \mathbb{R})$  and some gamma functions [12].
2. It remains to show that the integrals over  $z_i$  and  $z$  reproduce the said 6j symbol and gamma functions. This can be done by focusing (as a technical tool) on the expression for the JT disk four-point function, which is identical to (3.18) except that the four-level spectral density is replaced by the product of 4 disk spectral densities  $\rho_0(E_1) \dots \rho_0(E_4)$ .

The four-point function with a 6j symbol and gamma functions follows from various perspectives

on the quantization of JT gravity [50, 59, 62, 63]. The version with the  $z_i$  and  $z$  integrals follows from a direct quantization of the dual Schwarzian quantum mechanics [51, 64]. In particular, the propagator is the wavefunction  $\psi_{E_i}(z_i)$  multiplied by a “phase,” and those phases combine into (3.20). The factors  $e^{-\Delta_w \ell_w} e^{-\Delta \ell}$  are just the OTOC Schwarzian bilocals written out in this propagator language [51]. These two quantizations of the same theory prove the relation which we wanted.

Let us now label the four energies as

$$E_1 = E + \frac{\bar{\omega}}{4} + \frac{\omega_1}{2}, \quad E_2 = E - \frac{\bar{\omega}}{4} + \frac{\omega_2}{2}, \quad E_3 = E - \frac{\bar{\omega}}{4} - \frac{\omega_2}{2}, \quad E_4 = E + \frac{\bar{\omega}}{4} - \frac{\omega_1}{2}. \quad (3.22)$$

Here,  $\omega_1$  (resp.  $\omega_2$ ) is the energy difference between  $E_1$  and  $E_4$  (resp.  $E_2$  and  $E_3$ ), which will be exponentially small. On the other hand,  $\bar{\omega}$ , will be  $O(1)$  (details follow). Introducing an  $\alpha$  parameter as in (3.6), with

$$a = \frac{\beta}{2\pi} \alpha, \quad (3.23)$$

the Boltzmann weights in (3.18) combine into  $e^{-\beta E - a \bar{\omega}} e^{-i\omega_1 T_1 - i\omega_2 T_2}$ . We again work at fixed energy  $E$

$$\frac{\beta}{2\pi} = \frac{1}{2E^{1/2}}. \quad (3.24)$$

Below, this is always the meaning of  $\beta$ . Then we can define an angular variable at fixed energy, which we will again call  $\alpha$ , and its Legendre dual frequency  $\omega$  as follows

$$\alpha = 2E^{1/2} a, \quad \omega = \frac{\bar{\omega}}{2E^{1/2}}. \quad (3.25)$$

This leaves us with the rewriting of (3.18)

$$\begin{aligned} & G_{\Delta \Delta_w \text{ nonpert}}(T_1, T_2) \\ &= \int_{-\infty}^{+\infty} d\omega e^{-\alpha \omega} \int_{-\infty}^{+\infty} d\omega_1 e^{-i\omega_1 T_1} \int_{-\infty}^{+\infty} d\omega_2 e^{-i\omega_2 T_2} \frac{\rho(E_1, E_2, E_3, E_4)}{\rho_0(E_1) \dots \rho_0(E_4)} G_{\Delta \Delta_w \text{ disk}}(E_1, E_2, E_3, E_4), \end{aligned} \quad (3.26)$$

where

$$G_{\Delta \Delta_w \text{ disk}}(E_1, E_2, E_3, E_4) = \prod_{i=1}^4 \int_0^\infty \frac{dz_i}{z_i} \rho_0(E_i) \psi_{E_i}(z_i) \int_0^\infty \frac{dz}{z} I_3(z_1, z_2, z) I_3(z_3, z_4, z) e^{-\Delta_w \ell_w} e^{-\Delta \ell}. \quad (3.27)$$

We are interested in this triple Fourier transform, (3.26), in the *tau scaling limit*

$$T_1, T_2, e^{S_0} \rightarrow \infty, \quad T_1 e^{-S_0}, T_2 e^{-S_0} \text{ fixed}. \quad (3.28)$$

As explained around (2.18), this corresponds (as announced before) to considering  $\omega_1, \omega_2 \sim e^{-S_0}$ .



The integral (3.26) gets most of its meaningful contributions (because we consider  $\alpha \sim 1$ ) from  $\omega \sim 1$ . This means that for the overwhelming majority of the integrand we are considering  $(E_1, E_4) - (E_2, E_3) \gg e^{-S_0}$ . In other words, in the limit (3.28) one should consider  $E_1 - E_4 \sim e^{-S_0}$  and  $E_2 - E_3 \sim e^{-S_0}$ , but all other energy differences much *larger* than the inverse level spacing  $\sim e^{-S_0}$ . In this regime, the exact random matrix theory answer factorizes [39]

$$\frac{\rho(E_1 \dots E_4)}{\prod_{i=1}^4 \rho_0(E_i)} = \left(1 + \frac{\delta(\omega_1)}{\rho(E)} - \frac{\sin(\pi\rho(E)\omega_1)^2}{\pi\rho(E)^2\omega_1^2}\right) \left(1 + \frac{\delta(\omega_2)}{\rho(E)} - \frac{\sin(\pi\rho(E)\omega_2)^2}{\pi\rho(E)^2\omega_2^2}\right). \quad (3.29)$$

This corresponds to the statement that wormholes that “bridge over” the perturbation in (3.8) may be ignored in the tau scaling limit.<sup>22</sup> Note that in this regime we may treat the kernel as independent of  $\omega$ . Finally, to arrive at (3.7) we simply use as in section 2.4 the convolution theorem

$$\int_{-\infty}^{+\infty} d\omega_1 e^{i\omega_1 T_1} f(\omega_1) h(\omega_1) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} dT_{1\text{eff}} \int_{-\infty}^{+\infty} d\omega_a e^{i\omega_a (T_1 - T_{1\text{eff}})} f(\omega_a) \int_{-\infty}^{+\infty} d\omega_b e^{i\omega_b T_{1\text{eff}}} h(\omega_b). \quad (3.30)$$

Using (2.43) for the Fourier transform of (3.29) and applying the convolution theorem to (3.26) gives<sup>23</sup>

$$G_{\Delta\Delta_w \text{ nonpert}}(T_1, T_2) = \int_{-\infty}^{+\infty} dT_{1\text{eff}} \mathcal{F}(T_{1\text{eff}}|T_1) \int_{-\infty}^{+\infty} dT_{2\text{eff}} \mathcal{F}(T_{2\text{eff}}|T_2) G_{\Delta\Delta_w \text{ disk}}(T_{1\text{eff}}, T_{2\text{eff}}), \quad (3.31)$$

with

$$G_{\Delta\Delta_w \text{ disk}}(T_1, T_2) = \int_{-\infty}^{+\infty} d\omega e^{-\alpha\omega} \int_{-\infty}^{+\infty} d\omega_1 e^{-i\omega_1 T_1} \int_{-\infty}^{+\infty} d\omega_2 e^{-i\omega_2 T_2} G_{\Delta\Delta_w \text{ disk}}(E_1, E_2, E_3, E_4). \quad (3.32)$$

### 3.3 Semiclassical (scramblon) analysis of disk amplitude

We are interested in the semiclassical interpretation of the OTOC  $G_{\Delta\Delta_w \text{ disk}}(T_{1\text{eff}}, T_{2\text{eff}})$  and its decomposition (3.11). In this subsection, we will discuss exclusively the correlation function on the effective geometry, so we will drop the subscript “eff” from here on. We will follow closely [53], who adopt a very geometric way of thinking about the disk amplitude via “scramblons,”<sup>24</sup> The material of section 2.3

<sup>22</sup> More generally, the intuition is that wormholes over a line are only important when there is a large and opposite time on each side of the line.

<sup>23</sup> We could have used the convolution theorem also for the  $\omega$  dependence. However, even in our decomposition of the state (3.12) there is always the factor  $e^{-\Delta_w \ell_w}$ , which quickly decays to zero for large Lorentzian times. Then we are left with doing a short-Lorentzian time  $\omega$  integral of the four-level spectral density. This is exponentially dominated by the contributions without wormholes bridging over the blue line, (3.29), bringing us back to the situation we present here.

<sup>24</sup> Several other semiclassical descriptions of this OTOC are available [25, 53, 65], each with their own value.

prepares for the current discussion.<sup>25</sup> We start with repeating the disk four-point function (3.27)

$$\prod_{i=1}^4 \int_0^\infty \frac{dz_i}{z_i} \int_0^\infty dE_i \psi_{E_i}(z_i) e^{-N\beta_i E_i} \rho_0(E_i) \int_0^\infty \frac{dz}{z} I_3(z_1, z_2, z) I_3(z_3, z_4, z) e^{-\Delta_w \ell_w} e^{-\Delta \ell}. \quad (3.33)$$

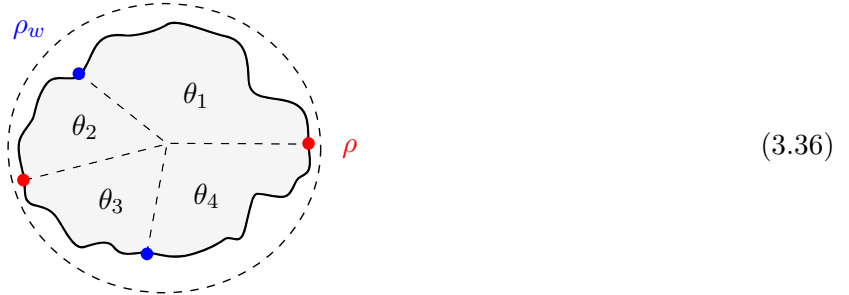
One can think of this as associated with a particle propagating near the boundary of AdS<sub>2</sub> [51]. The integration variables  $z_i$  and  $z$  parameterize the locations along the particle's trajectory on AdS<sub>2</sub> where operators are inserted. Using the 3 degrees of freedom of the SL(2, ℝ) isometry of AdS<sub>2</sub>, the 8 degrees of freedom parameterizing the locations of the 4 operators on AdS<sub>2</sub> reduce to 5 physical coordinates ( $z_i, z$ ). Parameterizing AdS<sub>2</sub> as

$$ds^2 = d\rho^2 + \sinh^2(\rho + 2 \log(\Phi_b/\varepsilon)) d\theta^2, \quad (3.34)$$

the renormalized geodesics length  $\ell_{ij}$  between the  $i$ -th and the  $j$ -th boundary locations is<sup>26</sup>

$$z_{ij} = e^{-\ell_{ij}/2} = 4 \frac{e^{-\frac{\rho_i + \rho_j}{2}}}{\sin \frac{\theta_i - \theta_j}{2}}. \quad (3.35)$$

A general off-shell contribution to the “particle propagator” (3.33) looks like [25, 51]



The sum of the angles spanned by the geodesics add up to  $2\pi$  (even off-shell)

$$\sum_{i=1}^4 \theta_i = 2\pi. \quad (3.37)$$

Combining 3 independent angles with the radial location  $\rho$  and  $\rho_w$  of the operator insertions  $\mathcal{O}_\Delta$  and  $\mathcal{O}_{\Delta_w}$  indeed gives 5 independent coordinates.<sup>27</sup>

Following [53], we introduced  $\Phi_b = N/2$  in (3.34), which makes it obvious how to take a classical limit

<sup>25</sup> The results of section 3.2 can in fact be reproduced (along the lines of section 2.3) by considering off-shell values of  $\alpha_i$  that correspond to the effective times by  $\alpha_i = 2\pi\beta_{i \text{ eff}}/\beta$ . We will not present this alternative derivation.

<sup>26</sup> It should be understood that  $0 < \theta_i - \theta_j < 2\pi$  such that the length  $\ell_{ij}$  is always positive.

<sup>27</sup> The SL(2, ℝ) isometry was used as follows. Two translations were used to place the two  $\mathcal{O}_\Delta$  at the same radial coordinate, *ditto* for the  $\mathcal{O}_{\Delta_w}$  insertions. The rotation then may be used to put the first insertion at  $\theta = 0$ .

of (3.33).<sup>28</sup> We have also rescaled energies with a factor  $N^2$  for transparency. Then labeling energies as in (3.22), imposing  $\omega, \omega_1, \omega_2 \ll E$ , and using the integral representations of the Bessel functions as in (2.25) and (2.26) (this introduces the  $\alpha_i$  variables), one finds the “action” for the amplitude (3.33)

$$\begin{aligned} & \prod_{i=1}^4 \int \frac{dz_i}{z_i} \frac{dz}{z} \psi_{E_i}(z_i) \rho_0(E_i) I_3(z_1, z_2, z) I_3(z_3, z_4, z) \\ & \stackrel{\text{class}}{=} \int d\lambda \prod_{i=1}^4 d\alpha_i d\theta_i d\rho d\rho_w \exp \left\{ N \left( E^{1/2} \sum_{i=1}^4 \alpha_i + \frac{\omega}{4} (\alpha_1 - \alpha_2 - \alpha_3 + \alpha_4) + \frac{\omega_1}{4E^{1/2}} (\alpha_1 - \alpha_4) \right. \right. \\ & \quad \left. \left. + \frac{\omega_2}{4E^{1/2}} (\alpha_2 - \alpha_3) + 4 \sum_{i=1}^4 e^{-\frac{\rho+\rho_w}{2}} \frac{\cos(\alpha_i/2)}{\sin(\theta_i/2)} - 2 \sum_{i=1}^4 \frac{e^{-\rho} + e^{-\rho_w}}{\tan(\theta_i/2)} + \lambda \left\{ \sum_{i=1}^4 \theta_i - 2\pi \right\} \right) \right\}. \end{aligned} \quad (3.38)$$

This is to be understood as an equality at the level of the action. We will not attempt to track one-loop factors (including integration measures and contours) in this illustrative calculation. Indeed, the final answer (3.53) of this calculation is well known (including the correct one-loop factors) [65]. We merely want to illustrate the geometric interpretation of the underlying calculation.

Because of the large factor  $N \gg 1$  up front in (3.38), one can do all of the integrals by saddle point. Keeping in mind the Boltzmann weight in (3.26), the  $\omega, \omega_1, \omega_2$  integrals localize on

$$4\alpha = \alpha_1 - \alpha_2 - \alpha_3 + \alpha_4, \quad 4iE^{1/2}T_1 = \alpha_1 - \alpha_4, \quad 4iE^{1/2}T_2 = \alpha_2 - \alpha_3, \quad (3.39)$$

leaving only the sum of  $\alpha_i$  unfixed. More generally the equations of motion are

$$\begin{aligned} 0 &= E_i^{1/2} - 2e^{-\frac{\rho+\rho_w}{2}} \frac{\sin(\alpha_i/2)}{\sin(\theta_i/2)}, \\ 0 &= \lambda - 2e^{-\frac{\rho+\rho_w}{2}} \frac{\cos(\alpha_i/2) \cos(\theta_i/2)}{\sin^2(\theta_i/2)} + \frac{e^{-\rho} + e^{-\rho_w}}{\sin^2(\theta_i/2)}, \\ 0 &= e^{-\frac{\rho}{2}} \sum_{i=1}^4 \frac{\cos(\alpha_i/2)}{\sin(\theta_i/2)} - e^{-\frac{\rho_w}{2}} \sum_{i=1}^4 \frac{1}{\tan(\theta_i/2)}, \\ 0 &= e^{-\frac{\rho_w}{2}} \sum_{i=1}^4 \frac{\cos(\alpha_i/2)}{\sin(\theta_i/2)} - e^{-\frac{\rho}{2}} \sum_{i=1}^4 \frac{1}{\tan(\theta_i/2)}, \\ 0 &= \sum_{i=1}^4 \theta_i - 2\pi. \end{aligned} \quad (3.40)$$

It is obvious from these equations that the unique classical solution is

$$\alpha_i = \theta_i, \quad \omega = \omega_1 = \omega_2 = 0, \quad e^{-\rho} = e^{-\rho_w} = \frac{E^{1/2}}{2}, \quad \lambda = -E^{1/2}. \quad (3.41)$$

---

<sup>28</sup> Here  $N$  plays the role of Newton’s constant  $G_N$  (hence large in the semiclassical limit). This is in addition to the small  $\varepsilon$  (the IR cutoff of AdS<sub>2</sub>) limit. Introducing  $N$  is not essential. Most JT references use  $N = 1$ , and we do this throughout this work, except in this section. The semiclassical limit is usually obtained by taking  $\beta_i$  small and shifting  $\rho$  according to (3.34) to make the action large. Indeed, JT amplitudes only depend on the ratio  $\beta_i/2\Phi_b$ , which in most early literature was called  $\beta_i/C$  [23, 25, 62, 65].

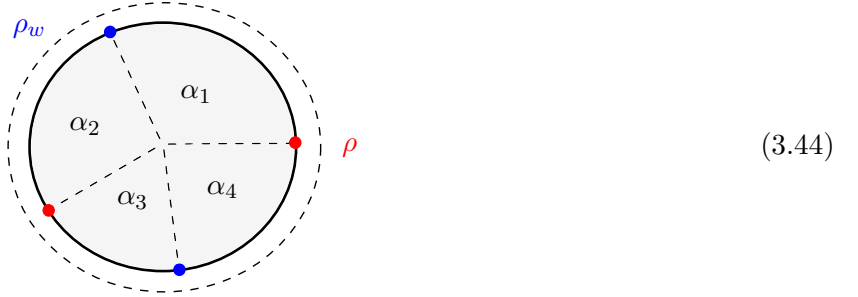
The first equation fixes the sum of the  $\alpha_i$  to  $2\pi$ , so combined with (3.39)

$$\alpha_1 = \frac{\pi}{2} + \alpha + i\frac{2\pi}{\beta}T_1, \quad \alpha_2 = \frac{\pi}{2} - \alpha + i\frac{2\pi}{\beta}T_2, \quad \alpha_3 = \frac{\pi}{2} - \alpha - i\frac{2\pi}{\beta}T_2, \quad \alpha_4 = \frac{\pi}{2} + \alpha - i\frac{2\pi}{\beta}T_1, \quad (3.42)$$

which can be summarized as the fact that  $\alpha_i$  are (on-shell) fractions of the boundary length (see (2.29))

$$\alpha_i = \frac{2\pi}{\beta} \beta_i. \quad (3.43)$$

The on-shell action only comes from the first term in the exponent of (3.38) and equals the entropy in JT gravity  $2\pi E^{1/2}$  (2.11). Geometrically, this saddle is obvious. The extremum of (3.36) is simply



$$(3.44)$$

The classical values  $\ell_{\text{bare}}$  and  $\ell_{w \text{ bare}}$  of the lengths of the geodesics between the two  $\mathcal{O}_\Delta$  and the two  $\mathcal{O}_{\Delta_w}$  operators, respectively, can be computed from this classical geometry as

$$e^{-\ell_{\text{bare}}/2} = \frac{2E^{1/2}}{\cosh(E^{1/2}(T_1 + T_2))}, \quad e^{-\ell_{w \text{ bare}}/2} = \frac{2E^{1/2}}{\cos(\alpha)}. \quad (3.45)$$

$\ell_{\text{bare}}$  matches with (2.8), up to an additive constant due to slightly different renormalization schemes.<sup>29</sup>

Naively the large  $N \gg 1$  factor in (3.38) guarantees a sharp saddle. However, it can occur that this saddle-point approximation breaks down [53]. This happens when  $e^{-2E^{1/2}T_i}N = O(1)$ , which is at the scrambling time [2]. In particular, it turns out that two of the dimensions in (3.38) do not have sharp extrema anymore in this case [53], so that one should do the honest integral over those soft directions.<sup>30</sup> In our setup the soft modes, or “scramblons” [25, 53, 66], capture Dray–’t Hooft shockwave interactions in the bulk. We will now see how this happens here.

On the classical saddle (3.42), both  $\alpha_i$  and  $\theta_i$  have large imaginary parts for large Lorentzian times. We are interested in configurations close to those saddles. For  $T_1$  and  $T_2$  to be positive and of order

<sup>29</sup>  $\ell_{w \text{ bare}}$  shows that one should regulate the  $\mathcal{O}_{\Delta_w}$  insertions with a Euclidean time separation  $\alpha \neq \pi/2$  as in footnote 20; otherwise,  $e^{-\Delta_w \ell_w}$  would blow up and backreact heavily on this classical geometry (the integrals in (3.33)). Also notice that  $e^{-\Delta_w \ell_{w \text{ bare}}}$  is *time-independent*, so it makes sense to divide it out as a normalization, as discussed around (3.12).

<sup>30</sup> This is familiar from spontaneous symmetry breaking. The soft modes are the pseudo Nambu–Goldstone modes.

the scrambling time or larger, the action (3.38) in this regime becomes

$$\exp \left\{ N \left( \sum_{i=1}^4 E_i^{1/2} \alpha_i + 4i \sum_{i=1}^2 e^{i \frac{\alpha_i - \theta_i + i\rho + i\rho_w}{2}} - 4i \sum_{i=3}^4 e^{-i \frac{\alpha_i - \theta_i - i\rho - i\rho_w}{2}} \right) + \text{suppressed} \right\}, \quad (3.46)$$

where the terms which we did not show would be at most of order  $e^{-2E^{1/2}T_i} N$  and so would *not* affect the saddle-point equations. This effective action only depends on the combination

$$i\theta_1 + \rho + \rho_w, \quad i\theta_4 - \rho - \rho_w \quad (3.47)$$

and ditto for  $\theta_2$  and  $\theta_3$ , respectively. This means that in this approximation  $\rho$  and  $\rho_w$  are not fixed, so that they remain as soft modes. The saddle for  $\alpha_i$  remains at (3.42), and for every values of  $\rho, \rho_w$  (the saddle-point manifold, or *approximate* saddle-point manifold to be precise) the  $\theta_i$  are pinned to

$$e^{i\frac{\theta_1}{2}} = e^{i\frac{\alpha_1}{2}} \frac{E^{1/2}}{2} e^{-\frac{\rho+\rho_w}{2}}, \quad e^{-i\frac{\theta_4}{2}} = e^{-i\frac{\alpha_4}{2}} \frac{E^{1/2}}{2} e^{-\frac{\rho+\rho_w}{2}}. \quad (3.48)$$

Evaluating the original action (3.38) (including Boltzmann weights) on these configurations, one ends up with the soft mode action with soft modes  $\rho$  and  $\rho_w$  to be integrated over<sup>31</sup>

$$\exp \left\{ N 2\pi E^{1/2} + \frac{N}{Z} \left( \frac{E^{1/2}}{2} e^\rho - 1 \right) \left( \frac{E^{1/2}}{2} e^{\rho_w} - 1 \right) \right\}, \quad (3.49)$$

with the ‘‘crossratio’’ [25, 65]

$$\frac{1}{Z} = 4E^{1/2} \cos(\alpha) e^{-2E^{1/2}|T_1|} + 4E^{1/2} \cos(\alpha) e^{-2E^{1/2}|T_2|}. \quad (3.50)$$

One can in fact check that this holds for all signs of  $T_1$  and  $T_2$ .

Following [25, 53], it is convenient to introduce<sup>32</sup>

$$x = \frac{E^{1/2}}{2} e^\rho - 1, \quad x_w = \frac{E^{1/2}}{2} e^{\rho_w} - 1. \quad (3.51)$$

For opposite signs of  $T_1$  and  $T_2$ , one finds

$$e^{-\ell/2} = \frac{e^{-\ell_{\text{bare}}/2}}{x+1}, \quad e^{-\ell_w/2} = \frac{e^{-\ell_w \text{ bare}/2}}{x_w+1}, \quad (3.52)$$

---

<sup>31</sup> The term  $N 2\pi E^{1/2}$  denotes the partition function at fixed energy and should be stripped off as part of the normalization, akin to the  $1/Z(\beta)$  in (2.10).

<sup>32</sup> From here on, we will work again with  $N = 1$  to be more consistent with our notation throughout the rest of the paper.

with the bare lengths given by the classical saddle (3.45). In the spirit of (3.11) one then writes

$$\begin{aligned}
G_{\Delta \Delta_w \text{ disk}}(T_1, T_2) &= \int_{-\infty}^{+\infty} d\ell \mathcal{A}_{\Delta_w \text{ disk}}(T_1, T_2, \ell) e^{-\Delta \ell} \\
&= e^{-\Delta \ell_{\text{bare}}} e^{-\Delta_w \ell_w \text{ bare}} \int dx dx_w e^{xx_w/Z} (x+1)^{-2\Delta} (x_w+1)^{-2\Delta_w} \\
&= \frac{e^{-\Delta \ell_{\text{bare}}} e^{-\Delta_w \ell_w \text{ bare}}}{\Gamma(2\Delta_w)} \int_0^\infty dx x^{2\Delta_w-1} e^{-x} (xZ+1)^{-2\Delta} \\
&= e^{-\Delta \ell_{\text{bare}}} e^{-\Delta_w \ell_w \text{ bare}} {}_2F_0(2\Delta, 2\Delta_w; ; -Z). \tag{3.53}
\end{aligned}$$

This reproduces equation (6.57) in [25], or (5.6) and (5.7) in [65].<sup>33</sup> To go from the second line to the third line we use the Hankel contour and the definition of the reciprocal Gamma function; for details see [67]. The contours and measure are only correct starting from the third line. It would be interesting to track the contours through the whole calculation. In the probe approximation and for large  $Z$ , this expression simplifies to

$$G_{\Delta \Delta_w \text{ disk}}(T_1, T_2) = e^{-\Delta \ell_{\text{bare}}} e^{-\Delta_w \ell_w \text{ bare}} Z^{-2\Delta}, \quad Z \gg 1, \quad T_1 T_2 < 0. \tag{3.54}$$

In the tau scaling limit  $|T_i| \rightarrow \infty$ , we are always in the regime  $Z \gg 1$  of this integral according to (3.50). This means that when the signs of  $T_1$  and  $T_2$  are opposite, a very strong shockwave interaction has been involved. Equation (3.54) reproduces the result (3.3) of the classical shockwave calculation of [45]<sup>34</sup>

$$\ell = 2E^{1/2}(|T_1| + |T_2|), \quad T_1 T_2 < 0, \tag{3.55}$$

It is reassuring to see this reappear from the exact JT gravity amplitude (3.33). Thus  $T_1 T_2 < 1$  involves large crossratios  $Z$  and is therefore preparing a dangerous slice.

For *equal* signs of  $T_1$  and  $T_2$  one finds that  $\ell$  is independent of  $x$ , with  $x$  defined in (3.51), whereas  $\ell_w$  is identical to (3.52). In this case, the  $x$  integral gives a delta function  $\delta(x_w)$ , such that both  $\ell$  and  $\ell_w$  localize onto their classical bare values (3.45). In other words, one finds

$$G_{\Delta \Delta_w \text{ disk}}(T_1, T_2) = e^{-\Delta \ell_{\text{bare}}} e^{-\Delta_w \ell_w \text{ bare}}, \quad T_1 T_2 > 0. \tag{3.56}$$

This is, therefore, the case in which shockwaves might as well have been ignored. The geometry of the  $\ell$  slice was not affected by the presence of the  $\mathcal{O}_{\Delta_w}$  particles. Again, this matches the discussion of [45], but now in the specific case of JT gravity.

<sup>33</sup> This hypergeometric function  ${}_2F_0(\cdot, \cdot; -Z)$  does not converge for all  $Z$ . It should be treated as the asymptotic series expansion of the Kummer  $U: Z^{-2\Delta} U(2\Delta, 1-2\Delta-2\Delta_w, 1/Z)$ . This is the form presented in most literature (such as [25,65]), but the symmetry between  $\Delta$  and  $\Delta_w$  is obscure in that form. The  ${}_2F_0$  function in (3.53) highlights the symmetry.

<sup>34</sup> This equation only concerns contributions which are exponentially large in  $e^{S_0}$ .

After normalization (see footnote 21), equations (3.54) and (3.56) reproduce the claimed result (3.14)

$$\mathcal{A}_{\Delta_w \text{ disk}}(T_1, T_2, \ell) = \delta(\ell - 2E^{1/2}(|T_1| + |T_2|)), \quad (3.57)$$

together (as explained) with the claimed classification of safe and dangerous slices (3.15)

$$\begin{aligned} T_1 T_2 < 0 \text{ or } |T_w| > |T| & \quad \text{dangerous} \\ T_1 T_2 > 0 \text{ or } |T_w| < |T| & \quad \text{safe} \end{aligned} \quad (3.58)$$

### 3.4 Conclusion

By now we have proven the steps leading up to equation (3.16) in section 3.1. As explained there, the probabilities that the dual bulk slice is safe and dangerous, respectively, are computed as

$$P_{\text{safe/danger}}(T_1, T_2) = \int_{\text{safe/danger}} dT_{1 \text{ eff}} dT_{2 \text{ eff}} \mathcal{F}(T_{1 \text{ eff}}|T_1) \mathcal{F}(T_{2 \text{ eff}}|T_2). \quad (3.59)$$

As shown in (3.58), the safe region corresponds to either both *effective* times positive  $T_{1 \text{ eff}}, T_{2 \text{ eff}} > 0$  or both times negative. The dangerous region is where  $T_{1 \text{ eff}}$  and  $T_{2 \text{ eff}}$  have opposite signs. This simple criterion allows us to decompose (3.59) into the “elementary” probabilities (2.44)

$$\boxed{P_{\text{safe}}(T_1, T_2) = P_{\text{exp}}(T_1)P_{\text{exp}}(T_2) + P_{\text{cont}}(T_1)P_{\text{cont}}(T_2)}, \quad (3.60)$$

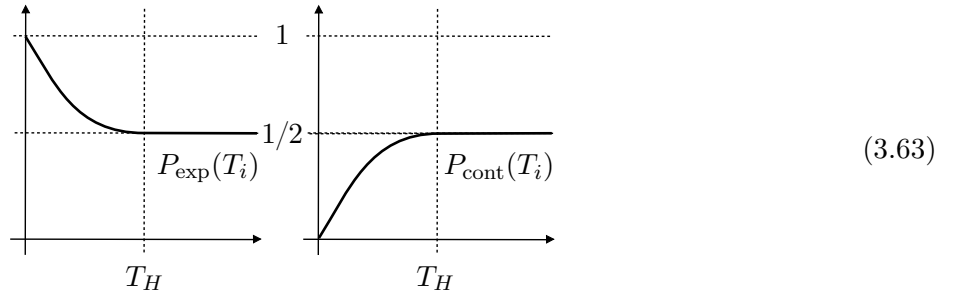
and similarly

$$P_{\text{danger}}(T_1, T_2) = P_{\text{exp}}(T_1)P_{\text{cont}}(T_2) + P_{\text{cont}}(T_1)P_{\text{exp}}(T_2). \quad (3.61)$$

Here, the probabilities for negative times should be understood as

$$P_{\text{exp}}(T_i < 0) = P_{\text{cont}}(-T_i > 0), \quad (3.62)$$

and vice versa. The probabilities  $P_{\text{exp}}(T_i)$  and  $P_{\text{cont}}(T_i)$  for  $T_i > 0$  are given by (2.46) and (2.47) with the “constant” subtracted, which we graphically reproduce here for reader’s convenience



Recalling that

$$T_1 = T - T_w, \quad T_2 = T + T_w, \quad (3.64)$$

this graphic immediately leads to a main conclusion of this section: for late enough times, the chances of encountering and not encountering a firewall (according to our definition) are fifty-fifty

$$\boxed{P_{\text{safe}}(T) = P_{\text{danger}}(T) = \frac{1}{2}}, \quad T > T_H + T_w. \quad (3.65)$$

To be more precise, at exponentially late times, there is a fifty-fifty chance that the dual slice contains a strongly back-reacting shockwave in the states we study in this section.

Another interesting regime is  $0 < T_2 \ll T_H$ , which corresponds to making a perturbation on the left, and having all the time evolution occur on the right. This might be closest in spirit to a setup obtained from collapse. One finds that the  $T_2$  slice is expanding ( $P_{\text{exp}}(T_2) = 1$ ,  $P_{\text{cont}}(T_2) = 0$ ), and therefore

$$P_{\text{safe}}(T_1, T_2) = P_{\text{exp}}(T_1), \quad P_{\text{danger}}(T_1, T_2) = P_{\text{cont}}(T_1). \quad (3.66)$$

This is thus a specific setup in which all expanding slices are safe and all contracting slices are dangerous, which is what [15] had in mind. Again, the odds of finding expanding (safe) and contracting (dangerous) slices asymptote to fifty-fifty at exponentially late times.

Before proceeding, we tie up one loose end. In the discussion above, we have ignored contributions from the second diagram in (3.8). This represents the (physically very real) possibility that the matter particle created by the insertion of  $\mathcal{O}_{\Delta_w}$  is carried away by a bra-ket wormhole in such a way that it bypasses (or avoids) the dual interior slice altogether. We call such events *avoided crossings*. Avoided crossings most definitely correspond to safe interior slices (as there are no particles, highly boosted or otherwise, in the slice altogether). Thus, these would contribute to  $P_{\text{safe}}(T)$  and, if dominant, might make the interior safe. However, as we show in appendix A, the opposite is actually true. As compared to the contributions discussed here, the avoided crossing is suppressed by a power  $1/T_H^2 \sim e^{-2S_0}$  and should be ignored in the tau-scaling limit (which involves  $T_H \rightarrow \infty$ ).

### 3.5 Effective time-folds

In light of the generalization in section 4, let us provide some additional intuition for the classification (3.58) of safe/dangerous slices. We can think of the correlator (3.5) as

$$\langle \Psi | \mathcal{O}_{\Delta_L} \mathcal{O}_{\Delta_R} | \Psi \rangle, \quad (3.67)$$

where the state being probed is the following evolution of the TFD

$$|\Psi\rangle = e^{-iH_L T_2} \mathcal{O}_{\Delta_w L} e^{-iH_L T_1} |\Psi_{\text{TFD}}\rangle. \quad (3.68)$$



We used  $H_R|\Psi_{\text{TFD}}\rangle = H_L|\Psi_{\text{TFD}}\rangle$  to put all Hamiltonian evolution on the left degrees of freedom. States of this type (and their dual) were analyzed in detail in [45], who more generally considered

$$|\Psi\rangle = e^{-iH_L T_{n+1}} \prod_{i=1}^n \left( \mathcal{O}_{\Delta_{w_i} L} e^{-iH_L T_i} \right) |\Psi_{\text{TFD}}\rangle. \quad (3.69)$$

Depending on the relative signs of  $T_1$  and  $T_2$  in (3.68) the boundary “time-fold” associated with (3.68) is time-ordered (TO) or “out-of-time-ordered” (OTO) [68, 69]. Following notation of [45], a fully time-ordered version of the time-fold preparing the state (3.69) is pictured as



$$\begin{array}{c} T_{n+1} \\ | \\ \bullet \\ \dots \\ | \\ \bullet \\ T_1 \end{array} \quad (3.70)$$

whereas generic time ordering (signs of  $T_i$ ) is pictured in [45] as a folded time contour



$$\begin{array}{c} T_{n+1} \\ | \\ \bullet \\ | \\ \dots \\ | \\ \bullet \\ T_1 \end{array} \quad (3.71)$$

The main focus of [45] was to discuss the complexity of these states or correspondingly [70] the length  $\ell$  of the dual bulk slice (focusing now on 2d gravity). The authors considered times much shorter than  $T_H$ , so wormholes did not play any role, and pictures such as (3.71) in this case represent physics occurring in the bulk at the classical level. We have seen, however, that wormholes change this. In particular, one can represent (3.12) and (3.13) using *effective time-folds*<sup>35</sup>

$$\langle \Psi | \dots | \Psi \rangle = \begin{array}{c} T_2 \\ | \\ \bullet \\ T_1 \end{array} + \text{wormholes} = \int_{-\infty}^{+\infty} dT_{1\text{eff}} \mathcal{F}(T_{1\text{eff}}|T_1) \int_{-\infty}^{+\infty} dT_{2\text{eff}} \mathcal{F}(T_{2\text{eff}}|T_2) \begin{array}{c} T_{2\text{eff}} \\ | \\ \bullet \\ T_{1\text{eff}} \end{array} \quad (3.72)$$

where we now take each contour to represent a *unique* bulk slice arising effectively through wormhole contributions. We emphasize that these contours do *not* represent boundary conditions in this equation! They are mnemonics that capture classical (effective) bulk slices.

<sup>35</sup> As a reminder, we only draw the Lorentzian time-fold for a ket. To compute the actual expectation value, we would need another copy of time-fold for the bra (along with the Euclidean circles). See, for example, equation (1.12) in [71] for a more complete drawing.

In the language developed above, the criterion (3.15) becomes (perhaps) more intuitive

$$\boxed{\begin{array}{ll} \text{effective OTO} & \text{dangerous} \\ \text{effective TO} & \text{safe} \end{array}}. \quad (3.73)$$

By “OTO” we mean that there is (at least) one switchback in the *effective* time contour that represents the dual semiclassical bulk slice. A “switchback” [45] is a fold in the time-fold. Thus, one way to phrase our findings is that wormholes may replace TO time-folds as boundary conditions with *effective* OTO time-folds. This results in dangerous shockwaves in unexpected places.

### Warning about nomenclature

We warn the reader that sometimes in the literature other criteria have been used to refer to a correlator as TO or OTO; here we have followed [45]. In particular, oftentimes one refers to the four-point function we considered in (3.5) as OTOC for generic (complex) times between all operators, simply because the operators are ordered as  $\mathcal{O}_\Delta \mathcal{O}_{\Delta w} \mathcal{O}_\Delta \mathcal{O}_{\Delta w}$  along the boundary. This is *not* the nomenclature we follow.

## 4 Exponentially dangerous states

In this section, we consider the logical generalization of section 3. What if we consider states prepared with multiple early perturbations? How dangerous are those?

We consider states obtained by perturbing the TFD on the left boundary at multiple times  $t = -T_{w_i}$  ( $i = 1, \dots, n$ ), where we take

$$T_{w_1} < \dots < T_{w_n}. \quad (4.1)$$

One could be interested in the case in which these perturbations are distributed in a “typical” manner. There are many ways to define typicality, as typicality refers to a choice of ensemble. In our setup, the most natural ensemble may be to consider early perturbations at times taking values on the whole real axis (within the recursion time). Then the vast majority of configurations will have (much) more than the Heisenberg time of separation between each particle. We show that such states almost always have firewalls, even though classically (ignoring wormholes) they would appear to be perfectly safe.

We do not want to argue that this ensemble is physically very relevant (which would mean it would be a good representative for black holes formed from gravitational collapse). Instead, we want to point out just quite *how* dramatic the situation can get once you properly account for wormholes.

## 4.1 Multiple shocks

We again consider the amplitude (3.67), but now it involved the state (3.69) with  $n$  perturbations

$$|\Psi\rangle = e^{-iH_L T_{n+1}} \prod_{i=1}^n \left( \mathcal{O}_{\Delta_{w_i} L} e^{-iH_L T_i} \right) |\Psi_{\text{TFD}}\rangle, \quad (4.2)$$

where

$$T_1 = T - T_{w_n}, \quad T_i = T_{w_{n+2-i}} - T_{w_{n+1-i}} \text{ for } i = 2 \dots n, \quad T_{n+1} = T_{w_1} + T. \quad (4.3)$$

The dual semiclassical geometry (ignoring wormholes) was discussed in [45]. Applied to 2d gravity, and for  $T_i$ 's much larger than the scrambling time, one finds a slice with  $n$  shockwaves and total length

$$\ell = 2E^{1/2} \sum_{i=1}^{n+1} |T_i|. \quad (4.4)$$

Equations (4.3) and (4.4) are the generalization of (3.3). The length (4.4) should be compared to the length of the dual slice in the *absence* of the shockwaves (or particle insertions)

$$\ell_{\text{bare}} = 2E^{1/2} |T_{\text{total}}|, \quad T_{\text{total}} = \sum_{i=1}^{n+1} T_i = 2T. \quad (4.5)$$

This implies that significant backreaction occurs as soon as *not all* the  $T_i$ 's have identical signs, or in other words as soon as *at least* one switchback has occurred in the notation of (3.71). The intuitive criterion (3.17) is thus the correct criterion in this more general case as well, in determining whether or not a bulk slice is dangerous or safe ( $\sim$  has a strong shock or not). We emphasize that time ordered (TO) in this context means that *all signs* are identical, i.e.  $\text{sgn}(T_1) = \dots = \text{sgn}(T_{n+1})$ .

We now claim that the precise generalization of (3.72) in the tau-scaling limit is

$$\langle \Psi | \dots | \Psi \rangle = \begin{array}{c} T_{n+1} \\ | \\ \bullet \\ \dots \\ | \\ T_1 \end{array} + \text{wormholes} = \prod_{i=1}^{n+1} \int_{-\infty}^{+\infty} dT_{i \text{ eff}} \mathcal{F}(T_{i \text{ eff}} | T_i) \begin{array}{c} T_{n+1 \text{ eff}} \\ | \\ \bullet \\ \dots \\ | \\ T_{1 \text{ eff}} \end{array}. \quad (4.6)$$

Here, we consider the generalized tau-scaling limit

$$T_1, T_2, \dots, T_{n+1}, e^{S_0} \rightarrow \infty, \quad T_1 e^{-S_0}, T_2 e^{-S_0}, \dots, T_{n+1} e^{-S_0} \text{ fixed}. \quad (4.7)$$

The derivation of (4.6) consists of several steps which are quite identical to those in section 3. Leaving the details to the interested reader, we simply summarize the main steps

1. In the generalization of (3.5), let us label the energies bordering boundary segments with Lorentzian time  $\pm T_i$  as  $E_{i \text{ bra}}$  and  $E_{i \text{ ket}}$ . Such regions are on opposite sides of the interior slice (red). Particle lines (blue) separate energies with different indices  $i$ . All regions get a Euclidean regulator. Then in our regime of interest the appropriate integration kernel (generalizing (3.29)) is

$$\frac{\rho(E_1 \dots E_{2n+2})}{\prod_{i=1}^{2n+2} \rho_0(E_i)} = \prod_{i=1}^{n+1} \left( 1 + \frac{\delta(\omega_i)}{\rho(E)} - \frac{\sin(\pi \rho(E) \omega_i)^2}{\pi \rho(E)^2 \omega_i^2} \right), \quad \omega_i = E_{i \text{ bra}} - E_{i \text{ ket}}. \quad (4.8)$$

The reason for this specific replacement is the Boltzmann weight  $\omega_i T_i$ , which at tau-scaling times favors bra and ket energies to be exponentially close. This forces us to consider bra-ket wormholes within each index  $i$ , such as those drawn in (3.5). On the other hand, there is no sense in which the integral prefers energies with different indices to be *exponentially* close together, since their energy differences have finite Euclidean Boltzmann weights (the aforementioned regulators). Therefore, other wormholes than those counted in (4.8) contribute negligibly.

2. With this result, the generalization of (3.31) is proven. Indeed, the convolution theorem can just be applied again. This suffices to prove (4.6). We would now like to go one step further and prove the generalization of (3.59). This is true *if* the disk amplitude decomposes similarly to (3.11)

$$G_{\Delta \Delta_{w_1} \dots \Delta_{w_n} \text{ disk}}(T_1, \dots, T_{n+1}) = \int_{-\infty}^{+\infty} d\ell \mathcal{A}_{\Delta_{w_1} \dots \Delta_{w_n} \text{ disk}}(T_1, \dots, T_{n+1}, \ell) e^{-\Delta \ell}, \quad (4.9)$$

with (in the semiclassical limit and after normalization)

$$\mathcal{A}_{\Delta_{w_1} \dots \Delta_{w_n} \text{ disk}}(T_1, \dots, T_{n+1}, \ell) = \delta(\ell - \ell(T_i)), \quad \ell(T_i) = 2E^{1/2} \sum_{i=1}^{n+1} |T_i|. \quad (4.10)$$

We did not check this explicitly for the multiple-shocks setup, but it is largely obvious. Equation (4.10) is identical to the statement that the two-point function in the shockwave geometry is classically  $e^{-\Delta \ell(T_i)}$ .<sup>36</sup> The least obvious part is the reasonable claim that the exact Schwarzian disk amplitude is dominated by that “classical” shockwave geometry. In section 3.3, we checked carefully that this is indeed the case, albeit only for the setup with one shock  $n = 1$ .

This results in the following generalization of (3.60)

$$\boxed{P_{\text{safe}}(T_1, \dots, T_{n+1}) = P_{\text{exp}}(T_1) \dots P_{\text{exp}}(T_{n+1}) + P_{\text{cont}}(T_1) \dots P_{\text{cont}}(T_{n+1})}. \quad (4.11)$$

This counts only those contributions where *all* signs of  $T_{i \text{ eff}}$  are equal. Any other combination of signs

---

<sup>36</sup> We normalize the amplitude with that for fixed boundary times, not effective times. One might worry that dependence on effective times might creep in via the factor  $e^{-\Delta_i \ell_{w_i}(T_{j \text{ eff}})}$ . This does not happen. Indeed, the perturbation particles on-shell follow geodesics in an unperturbed TFD at  $t = 0$ , because the “total time” left and right of them adds up to zero, and therefore the length  $\ell_{w_i}$  is in fact independent of time (and hence also of effective times). So there is no effective time dependence coming in via the normalization, at least classically. This statement is the generalization of footnote 21.

has at least one switchback, and should be considered dangerous

$$P_{\text{danger}}(T_1, \dots, T_{n+1}) = \sum_{\text{signs unequal}} P_{\text{exp}}(\pm T_1) \dots P_{\text{exp}}(\pm T_{n+1}), \quad (4.12)$$

with probabilities at negative times computed as in (3.62). We will now analyze these probabilities in various scenarios, depending on the parametric choices of boundary times  $T_i$ .

## 4.2 Firewall probabilities

The clearest physical picture is when one considers either  $T_i \ll T_H$  or  $T_i \gg T_H$ . In the former case, one recovers classical physics, the geometry simply expands

$$P_{\text{exp}}(T_i) = 1, \quad T_i \ll T_H. \quad (4.13)$$

In the latter case, we probe the plateau region where the chances of expanding and contracting branches are equal

$$P_{\text{exp}}(T_i) = \frac{1}{2}, \quad T_i \gg T_H. \quad (4.14)$$

To demonstrate our methods, we will explore three different scenarios.

1. The  $n$  particles are all separated by  $T_i \gg T_H$ . In this case, all the factors in (4.11) have reached their probability plateaus (4.14). Thus, for extremely late times  $T > T_{w_n} + T_H$ , we have

$$P_{\text{safe}}(T) = \frac{2}{2^{n+1}} = \frac{1}{2^n}. \quad (4.15)$$

Similarly, by counting all the dangerous permutations in (4.12) we obtain

$$\boxed{P_{\text{danger}}(T) = 1 - \frac{1}{2^n}}, \quad T > T_{w_n} + T_H. \quad (4.16)$$

This is the result announced in the introduction (1.10), and one of the main points of our paper. For late times, these states (which as pointed out at the beginning of this section, are typical in some sense) *almost certainly* have firewalls if we consider many perturbations  $n$ .

2. The particles are bunched together around  $t = 0$  at timescales much shorter than the Heisenberg time, but longer than the scrambling time  $T_H \gg T_2, \dots, T_n \gg T_S = (\beta/2\pi) \log S_0$ . In this case, the  $n - 1$  intermediate times give purely expanding probabilities (4.13). This results in

$$P_{\text{safe}}(T_1, \dots, T_{n+1}) = P_{\text{exp}}(T_1) P_{\text{exp}}(T_{n+1}) \quad (4.17)$$

and

$$P_{\text{danger}}(T_1, \dots, T_{n+1}) = P_{\text{cont}}(T_1)P_{\text{cont}}(T_{n+1}) + P_{\text{exp}}(T_1)P_{\text{cont}}(T_{n+1}) + P_{\text{cont}}(T_1)P_{\text{exp}}(T_{n+1}). \quad (4.18)$$

This results for  $T > T_H$  in the plateau

$$\boxed{P_{\text{danger}}(T) = \frac{3}{4}}, \quad T > T_H. \quad (4.19)$$

It may look puzzling why the contracting-contracting term is dangerous if we compare with the  $n = 1$  case in section 3 (see (3.61)). The reason is that the ordering of the  $n > 1$  operator insertions already picks a “preferred” time axis. Indeed, the contracting-contracting case has a non-trivial time-fold [45]:



Thus the dual geometry still contains a shock. While this shock is “less severe” than in the other cases, it *is* dangerous. Indeed, the effect on the geodesic  $\ell$ , whilst much less than  $e^{S_0}$ , is still much larger than  $\log S_0$ .

If the perturbations are bunched around  $T_w \gg T_H$ , we reach the same asymptotics for  $T > T_H + T_w$ .

3. If successive particles are separated by timescales less than the scrambling time (such as thermal) one may essentially treat them as one particle in the switch-back diagrams of [45]. Therefore, in a scenario where all the particles are bunched around  $t = 0$  and  $T_2, \dots, T_n \ll T_S$ , one finds

$$\boxed{P_{\text{danger}}(T) = \frac{1}{2}}, \quad T > T_H. \quad (4.21)$$

Indeed, in this case the contracting-contracting configuration (4.20) has essentially no switch-back; in other words, the shockwave does not severely backreact on  $\ell$  [45]. Therefore, in this case the contracting-contracting term should be considered to contribute to  $P_{\text{safe}}(T)$ .

In summary, black holes created with early perturbations on the other side of the TFD are generically very dangerous at post-Heisenberg times. Depending on the detailed setup and one’s notion of typicality, at the very best the probability of a firewall is  $1/2$ . At worst, firewalls are *guaranteed*.

## 5 Concluding remarks

To end this paper, we will point out that our results (using as an example (2.45)) can be obtained by summing (following [37, 38, 46]) a perturbatively convergent series (in genus) of wormhole amplitudes, and we will also identify the corresponding Lorentzian wormhole geometries [47, 48]. This is discussed, respectively, in sections 5.2 and 5.3.

Before doing so, however, we propose several ways to potentially improve on our setup. Our general attitude is that despite some shortcomings in our setup, our results seem physically insightful. We think that our techniques and ideas will also help in attacking the improvements that we propose below.

### 5.1 Room for improvement

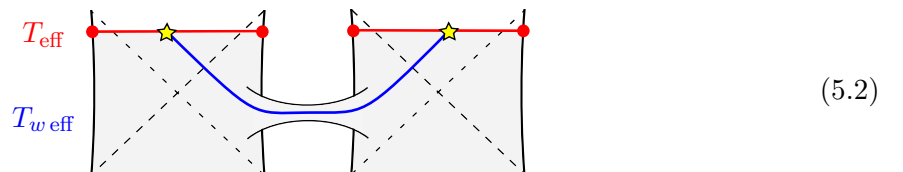
Here we list potential points of critique on our setup and how to improve on them.

1. We are not describing measurements performed by an infalling observer in quantum gravity, as it is not known how to describe such infalling observers in quantum mechanics. (For some interesting recent progress on this problem see for instance [60, 61].) It might be that our physical conclusion changes a lot when an observer is included. As a very crude approximation, one could contemplate for instance modeling an observer by a particle with  $\Delta \rightarrow 0$  that stretches the interior slice. More precisely, if we take the  $\Delta \rightarrow 0$  limit while keeping  $\Delta \gg 1/T_H \gg 1/\Lambda$  with  $\Lambda$  an IR cutoff for  $T_{\text{eff}}$ , then the exponential suppression  $e^{-\Delta \ell(T_{\text{eff}})}$  ensures essentially that wormhole corrections in (2.45) do not contribute

$$P_{\text{cont } \Delta}(T) = \lim_{\Delta \rightarrow 0} \int_{-\Lambda}^0 dT_{\text{eff}} \mathcal{F}(T_{\text{eff}}|T) e^{-2\Delta\sqrt{E}|T_{\text{eff}}|} = 0, \quad P_{\text{exp } \Delta}(T) = 1. \quad (5.1)$$

Here we considered  $\Delta \ll 1/T$  such that  $e^{-2\Delta\sqrt{E}T} \rightarrow 1$ . This is not a realistic model; but it shows the potentially far-reaching consequences of carefully defining the observer.

2. We considered probe matter perturbations, *not* dynamical QFT with particle-antiparticle pairs. A consequence of this is that in our setup and that of [15], the pure TFD is *obviously* safe. However, when particle-antiparticle pairs can be created, it is fathomable that wormholes could produce dangerous shocks in unexpected places. We believe such effects would be closer in spirit with the original firewall ideas [16–18]. For instance, one might imagine (schematically) a surprising shock even in the pure TFD, due to a process like:



Unfortunately studying dynamical matter on wormhole geometries is challenging. Indeed, in JT gravity it leads to UV divergences. To make further progress on this question, it seems one would have to first resolve that issue. One could start with studying one particle loop. Alternatively, one could consider a UV regulated q-deformation of JT gravity [43], which has a bulk interpretation as a different simple 2d dilaton gravity [72,73]. In that theory matter loops seem better behaved. Supersymmetric JT may also deal with these UV divergences [44].

3. As discussed in section 2.5, we renormalized our amplitudes by subtracting off an infinite constant. We believe this is physically well motivated, following identical logic in [10]. Nevertheless, this is a subtle point, one that for instance Stanford and Yang [15] put a lot of effort in trying to avoid.<sup>37</sup> Obviously in a UV complete theory such subtractions should not be required. From this point of view, it may be worthwhile trying to reformulate these types of questions in the matrix integral dual [8] of JT gravity, and in particular in one member of the ensemble [58,74–78] (even though individual members of the ensemble may be dangerous for very different reasons [79]). Progress on this front is being made [49].
4. Orthogonal to the previous problems, it would be an improvement to mimic our setup for black holes in the sky (formed from a gravitational collapse), which are not a TFD with certain amount of perturbations on the left (which we studied). A first step in this direction would be to consider pure states in AdS. This should be possible, probably using end-of-the-world (EOW) branes in JT gravity [80,81]. What is the physically meaningful question to ask in an EOW brane setup? Naively, the slice is always dangerous when it ends on the EOW brane (spacetime ending is quite dramatic), but this conclusion is likely overly simplistic. A second step is to allow the black hole to evaporate, perhaps along the lines of [13]. More realistic models suffer often from the lack of a simple exact Euclidean path integral description, and seem more difficult to study using our techniques.<sup>38</sup>

## 5.2 Perturbative firewall probability plateaus

We remind the reader of our semiclassical answer for the transition kernel (2.36)

$$\mathcal{F}(T_{\text{eff}}|T) = \frac{1}{2\pi\rho(E)^2} \int_{-\infty}^{+\infty} d\omega e^{i\omega(T_{\text{eff}}-T)} \left( \rho(E)^2 + \delta(\omega)\rho(E) - \frac{\sin(\pi\rho(E)\omega)^2}{\pi\omega^2} \right). \quad (5.3)$$

This can be rewritten as

$$\mathcal{F}_{\text{un-norm}}(T_{\text{eff}}|T) = \delta(T - T_{\text{eff}}) \rho(E)$$

---

<sup>37</sup> We are not sure that factorizing the empty partition function as Stanford and Yang do is the best setup to improve on this. Physical observations involve measurements (as in our setup), while the empty partition function does not.

<sup>38</sup> See for instance [29–34] for more discussion on black holes formed from collapse.



$$+ \frac{1}{2\pi\rho(E)} \frac{1}{i\pi} \int_{-i\infty}^{+i\infty} d\beta e^{2\beta E} Z_{\text{conn}}(\beta + i(T - T_{\text{eff}}), \beta - i(T - T_{\text{eff}})). \quad (5.4)$$

Here, we have multiplied  $\mathcal{F}(T_{\text{eff}}|T)$  by  $\rho(E)$ , as we want to compare it with an un-normalized sum over geometries in the gravitational path integral. Furthermore we introduced

$$Z_{\text{conn}}(\beta + iT, \beta - iT) = \int_0^\infty dE e^{-2\beta E} \min(|T|/2\pi, \rho(E)). \quad (5.5)$$

One of the main points of [37, 38, 46] was that this equation for the connected two-boundary amplitude is *exact* for dilaton gravity (in the tau-scaling limit), and that it admits a Taylor series in  $T$

$$Z_{\text{conn}}(\beta + iT, \beta - iT) = \frac{T}{4\pi\beta} + \sum_{g=1}^{\infty} P_{g-1}(\beta) T^{2g+1}, \quad (5.6)$$

where the polynomial  $P_{g-1}(\beta)$  is computed by a contour integral around the real axis

$$P_{g-1}(\beta) = -\frac{1}{(2\pi)^{2g+1}(2g)(2g+1)} \oint_R dE \rho(E)^{-2g} e^{-2\beta E}. \quad (5.7)$$

Because  $\rho(E) \sim e^{S_0}$ , this Taylor series is actually the gravitational genus expansion, with  $g$  the number of handles (or wormholes). This series is reproduced by the Weil-Peterson polynomials in (2.14) [37, 38, 46]. So, the sine kernel (2.19) of random matrix theory in the time domain (5.5) is *perturbatively* (in  $g$ ) accessible in gravity (in the tau-scaling limit).<sup>39</sup>

Here we want to point out that the same is true for our un-normalized probability (5.4). In particular, at fixed temperature and following the same steps resulting in equation (A.14) or (A.19) in [47], one obtains the genus expansion<sup>40</sup>

$$\begin{aligned} \mathcal{F}_{\text{un-norm}}(T_{\text{eff}}|T) &= \delta(T - T_{\text{eff}}) Z(\beta) + \frac{1}{4\pi^2} \int_0^\infty dE e^{-\beta E} \rho(E)^{-1} |T - T_{\text{eff}}| \\ &\quad - \sum_{g=1}^{\infty} \frac{1}{2g(2g+1)(2\pi)^{2g+2}} \oint_R dE e^{-\beta E} \rho(E)^{-1-2g} |T - T_{\text{eff}}|^{2g+1}. \end{aligned} \quad (5.8)$$

This sum is convergent [37, 38], so the firewall probability plateau in (1.6) is perturbatively accessible.

### 5.3 Lorentzian spacetimes

We now wonder how to obtain (5.8) from purely Lorentzian geometries. Besides an independent interest in Lorentzian wormhole geometries, this is relevant to us as it provides another indication that effective

<sup>39</sup> Backing off from this tau-scaling limit, one does have to consider non-perturbative effects [8, 82–86].

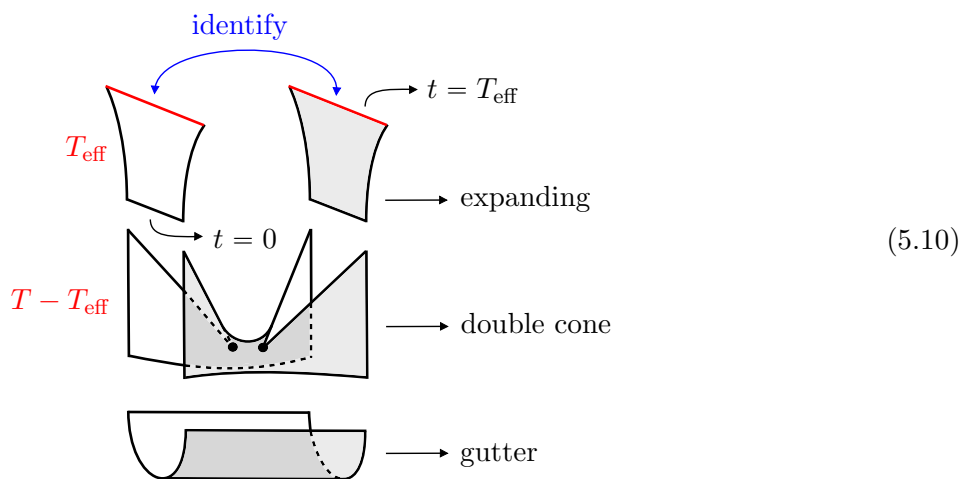
<sup>40</sup> One can use this to compute the two-point function (2.10) in the tau scaling limit by inserting  $e^{-\Delta\ell(T_{\text{eff}})}$  (taken from equation (2.8)). However, in the tau-scaling limit this exponential backreacts heavily, since  $\ell$  is exponentially large in the entropy. This thus projects onto  $T_{\text{eff}} = 0$ . The equation for the two-point function then reduces to equation (4.5) in [47].

times  $T_{i\text{eff}}$  determine the true Lorentzian slice of spacetime that is being probed. This interpretation was motivated more in sections 2.1 and 2.3.

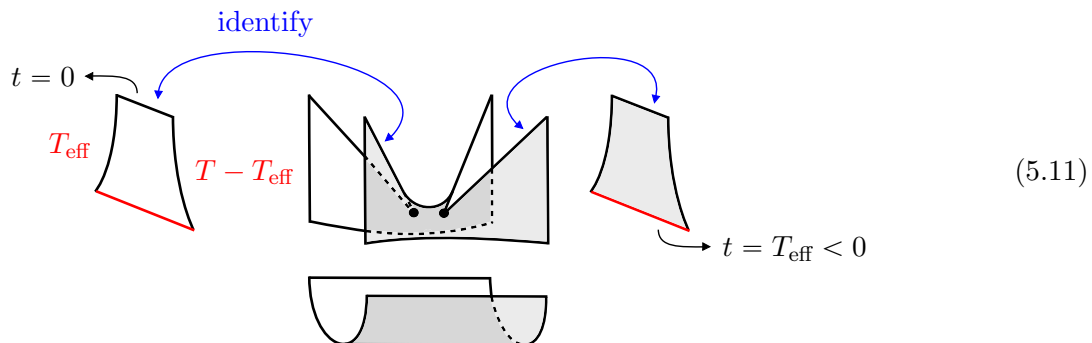
The relevant Lorentzian spacetimes are a mild modification of those discussed in section 4 of [47], which we follow closely, and to which we refer readers for a more pedagogical explanation. Like in [47], we do not have enough control of Lorentzian JT gravity to reproduce the full details of (5.8). Instead, we *can* reproduce the semiclassical ( $\sim$  large energy) approximation<sup>41</sup>

$$\mathcal{F}_{\text{un-norm}}(T_{\text{eff}}|T) \sim \int_{\Lambda_g}^{\infty} dE e^{-\beta E} \rho(E)^{-1-2g} |T - T_{\text{eff}}|^{2g+1}. \quad (5.9)$$

This is reproduced, for  $T_{\text{eff}} > 0$  (expanding) and  $g = 0$ , by the following geometries



whereas for  $T_{\text{eff}} < 0$  (contracting) and  $g = 0$ , the relevant (mostly) Lorentzian spacetimes are



These are to be compared with equation (4.16) in [47]. The “gutter” is half of the Euclidean wormhole, which is glued onto patches of the double-cone spacetime [6] in the way indicated (see (4.15) in [47] for more details). The black dots denote specific curvature singularity called “crotches” [47,87]. The point is that part of the time evolution imposed by the boundary conditions can be “absorbed” by a portion

<sup>41</sup> One could say that this is a poor approximation to (5.8), which obtains its main contributions from very low energies [37]. Classical physics is a poor approximation for low energies. Thus, semiclassically, the best one could hope for might be (5.9).

of double-cone spacetime. Any time slice of that double cone is identical to the global TFD at  $t = 0$ , with metric and dilaton

$$ds = d\rho = \frac{d\sigma}{\sin(\sigma)}, \quad \Phi = E^{1/2} \cosh(\rho) = \frac{E^{1/2}}{\sin(\sigma)}, \quad (5.12)$$

which means we can glue this smoothly to a  $t = 0$  TFD, and use the remainder of the boundary time evolution to expand that slice either into the future or past (depending on whether this remaining time  $T_{\text{eff}}$  is positive or negative). For  $T - T_{\text{eff}} < 0$  one simply lets the double-cone evolve to the past.

A factor  $|T - T_{\text{eff}}|$  comes from the twist zero mode of the double cone [47]. For  $g > 0$  the spacetimes are those in (5.10) and (5.11) with additional crotch singularities inserted at mirrored locations on the double-cone pieces of the  $g = 0$  spacetimes, in analogy to equation (4.18) in [47]. Their time coordinates on the double cone are zero modes and explain the additional powers of  $|T - T_{\text{eff}}|$  in (5.9).<sup>42</sup> The power of  $\rho(E)$  comes from the on-shell actions of the crotches [47]. So, summing over classical Lorentzian wormhole geometries, following the rules put forward in [47], reproduces (5.9). And indeed, the slice in which the measurement takes place has the geometry of the usual TFD, but at *effective* time  $t = T_{\text{eff}}$ .

## Acknowledgments

We thank Jorrit Kruthoff, Geoff Penington, Douglas Stanford, Mykhaylo Usatyuk, Shunyu Yao, and Ying Zhao for useful discussions. AB was funded by ERC-COG Grant NP-QFT No. 864583 and by INFN Iniziativa Specifica GAST. The work of CHC and YN was supported in part by the Department of Energy, Office of Science, Office of High Energy Physics under QuantISED award DE-SC0019380 and contract DE-AC02-05CH11231. The work of CHC was also supported in part by the Department of Energy through DE-FOA-0002563 and by AFOSR award FA9550-22-1-0098. The work of YN was also supported in part by MEXT KAKENHI grant number JP20H05850, JP20H05860.

## A Avoided crossings

Here we consider the avoided crossing

$$G_{\Delta \Delta_w \text{ nonpert}}(T_1, T_2) \supset \text{Diagram} \quad (A.1)$$

<sup>42</sup> The saddle-point equations localize the additional crotches on the double cone piece. Indeed, there is no saddle on the expanding piece. This explains the powers of  $|T - T_{\text{eff}}|$  as opposed to simple powers of  $T$ .

The exact amplitude is

$$\frac{e^{-2S_0}}{Z(\beta)} \int dE e^{-\beta E} \rho(E) \int_{-\infty}^{+\infty} d\ell \psi_E(\ell) \psi_E(\ell) e^{-\Delta\ell} \int_{-\infty}^{+\infty} d\ell_w \psi_E(\ell_w) \psi_E(\ell_w) e^{-\Delta_w \ell_w}. \quad (\text{A.2})$$

At fixed energy, this simply factorizes

$$e^{-2S_0} \int_{-\infty}^{+\infty} d\ell \psi_E(\ell) \psi_E(\ell) e^{-\Delta\ell} \int_{-\infty}^{+\infty} d\ell_w \psi_E(\ell_w) \psi_E(\ell_w) e^{-\Delta_w \ell_w}. \quad (\text{A.3})$$

Each of these integrals computes the expectation value of a two-point function in the TFD at  $t = 0$ , up to a factor of  $1/\rho_0(E)$ . This is clear from (2.10). Alternatively, one can just do the  $\ell$  and  $\ell_w$  integrals, which results in gamma functions [50]. Then, using Stirling's approximation for the gamma functions, one indeed recovers

$$G_{\Delta \Delta_w \text{ avoided}}(T_1, T_2) = \frac{1}{T_H^2} e^{\Delta \log(E)} e^{\Delta_w \log(E)}, \quad T_H = 2\pi\rho(E). \quad (\text{A.4})$$

Stripping off the normalization factor  $e^{\Delta_w \log(E)}$  results in

$$\mathcal{A}_{\text{avoided}}(T_1, T_2, \ell) = \frac{1}{T_H^2} \delta(\ell + \log(E)), \quad (\text{A.5})$$

therefore giving the contribution

$$P_{\text{safe}}(T_1, T_2) \supset P_{\text{avoided}}(T_1, T_2) = \frac{1}{T_H^2}. \quad (\text{A.6})$$

This is negligible in the tau-scaling limit. One should compare (A.5) with our main contribution to the amplitude due to wormholes that we studied in the main text (2.45). That amplitude too is suppressed by  $1/T_H^2$ . However, it has support for large ranges of  $T_{\text{eff}}$  of order  $T_H$ , and the integral over such a large range results in a leading order correction to the safe/dangerous probabilities, unlike (A.5) which has delta support on  $T_{\text{eff}} = 0$  resulting in a subleading contribution (A.6).

## References

- [1] J. M. Maldacena, “The Large N limit of superconformal field theories and supergravity,” *Adv. Theor. Math. Phys.* **2** (1998) 231–252, [arXiv:hep-th/9711200](#).
- [2] S. H. Shenker and D. Stanford, “Black holes and the butterfly effect,” *JHEP* **03** (2014) 067, [arXiv:1306.0622 \[hep-th\]](#).
- [3] J. S. Cotler, G. Gur-Ari, M. Hanada, J. Polchinski, P. Saad, S. H. Shenker, D. Stanford, A. Streicher, and M. Tezuka, “Black holes and random matrices,” *JHEP* **05** (2017) 118,

- [arXiv:1611.04650](#) [hep-th].
- [4] J. M. Maldacena, “Eternal black holes in anti-de Sitter,” *JHEP* **04** (2003) 021, [arXiv:hep-th/0106112](#).
- [5] D. N. Page, “Average entropy of a subsystem,” *Phys. Rev. Lett.* **71** (1993) 1291–1294, [arXiv:gr-qc/9305007](#).
- [6] P. Saad, S. H. Shenker, and D. Stanford, “A semiclassical ramp in SYK and in gravity,” [arXiv:1806.06840](#) [hep-th].
- [7] P. Saad, “Late time correlation functions, baby universes, and ETH in JT gravity,” [arXiv:1910.10311](#) [hep-th].
- [8] P. Saad, S. H. Shenker, and D. Stanford, “JT gravity as a matrix integral,” [arXiv:1903.11115](#) [hep-th].
- [9] A. Blommaert, T. G. Mertens, and H. Verschelde, “Clocks and rods in Jackiw-Teitelboim quantum gravity,” *JHEP* **09** (2019) 060, [arXiv:1902.11194](#) [hep-th].
- [10] L. V. Iliesiu, M. Mezei, and G. Sárosi, “The volume of the black hole interior at late times,” *JHEP* **07** (2022) 073, [arXiv:2107.06286](#) [hep-th].
- [11] J. Kruthoff, “Higher spin JT gravity and a matrix model dual,” *JHEP* **09** (2022) 017, [arXiv:2204.09685](#) [hep-th].
- [12] A. Blommaert, “Dissecting the ensemble in JT gravity,” *JHEP* **09** (2022) 075, [arXiv:2006.13971](#) [hep-th].
- [13] G. Penington, S. H. Shenker, D. Stanford, and Z. Yang, “Replica wormholes and the black hole interior,” *JHEP* **03** (2022) 205, [arXiv:1911.11977](#) [hep-th].
- [14] A. Almheiri, T. Hartman, J. Maldacena, E. Shaghoulian, and A. Tajdini, “Replica wormholes and the entropy of Hawking radiation,” *JHEP* **05** (2020) 013, [arXiv:1911.12333](#) [hep-th].
- [15] D. Stanford and Z. Yang, “Firewalls from wormholes,” [arXiv:2208.01625](#) [hep-th].
- [16] A. Almheiri, D. Marolf, J. Polchinski, and J. Sully, “Black holes: complementarity or firewalls?,” *JHEP* **02** (2013) 062, [arXiv:1207.3123](#) [hep-th].
- [17] A. Almheiri, D. Marolf, J. Polchinski, D. Stanford, and J. Sully, “An apologia for firewalls,” *JHEP* **09** (2013) 018, [arXiv:1304.6483](#) [hep-th].
- [18] D. Marolf and J. Polchinski, “Gauge/gravity duality and the black hole interior,” *Phys. Rev. Lett.* **111** (2013) 171301, [arXiv:1307.4706](#) [hep-th].

- [19] J. Maldacena and L. Susskind, “Cool horizons for entangled black holes,” *Fortsch. Phys.* **61** (2013) 781–811, [arXiv:1306.0533 \[hep-th\]](#).
- [20] F. Haake, S. Gnutzmann, and M. Kuś, *Quantum Signatures of Chaos; 4th ed.* Springer series in synergetics. Springer, Dordrecht, 2018.
- [21] R. Jackiw, “Lower dimensional gravity,” *Nuclear Physics B* **252** (1985) 343–356.
- [22] C. Teitelboim, “Gravitation and hamiltonian structure in two spacetime dimensions,” *Physics Letters B* **126** no. 1-2, (1983) 41–45.
- [23] J. Engelsöy, T. G. Mertens, and H. Verlinde, “An investigation of AdS<sub>2</sub> backreaction and holography,” *JHEP* **07** (2016) 139, [arXiv:1606.03438 \[hep-th\]](#).
- [24] K. Jensen, “Chaos in AdS<sub>2</sub> holography,” *Phys. Rev. Lett.* **117** no. 11, (2016) 111601, [arXiv:1605.06098 \[hep-th\]](#).
- [25] J. Maldacena, D. Stanford, and Z. Yang, “Conformal symmetry and its breaking in two dimensional Nearly Anti-de-Sitter space,” *PTEP* **2016** no. 12, (2016) 12C104, [arXiv:1606.01857 \[hep-th\]](#).
- [26] T. G. Mertens and G. J. Turiaci, “Solvable models of quantum black holes: a review on Jackiw–Teitelboim gravity,” *Living Rev. Rel.* **26** no. 1, (2023) 4, [arXiv:2210.10846 \[hep-th\]](#).
- [27] S. H. Shenker and D. Stanford, “Multiple shocks,” *JHEP* **12** (2014) 046, [arXiv:1312.3296 \[hep-th\]](#).
- [28] L. Susskind, “Black Holes at Exp-time,” [arXiv:2006.01280 \[hep-th\]](#).
- [29] Y. Nomura, “Reanalyzing an evaporating black hole,” *Phys. Rev. D* **99** no. 8, (2019) 086004, [arXiv:1810.09453 \[hep-th\]](#).
- [30] Y. Nomura, “Spacetime and universal soft modes — black holes and beyond,” *Phys. Rev. D* **101** no. 6, (2020) 066024, [arXiv:1908.05728 \[hep-th\]](#).
- [31] Y. Nomura, “Interior of a unitarily evaporating black hole,” *Phys. Rev. D* **102** no. 2, (2020) 026001, [arXiv:1911.13120 \[hep-th\]](#).
- [32] K. Langhoff and Y. Nomura, “Ensemble from coarse graining: reconstructing the interior of an evaporating black hole,” *Phys. Rev. D* **102** no. 8, (2020) 086021, [arXiv:2008.04202 \[hep-th\]](#).
- [33] Y. Nomura, “Black hole interior in unitary gauge construction,” *Phys. Rev. D* **103** no. 6, (2021) 066011, [arXiv:2010.15827 \[hep-th\]](#).
- [34] C. Murdia, Y. Nomura, and K. Ritchie, “Black hole and de Sitter microstructures from a semiclassical perspective,” *Phys. Rev. D* **107** no. 2, (2023) 026016, [arXiv:2207.01625 \[hep-th\]](#).

- [35] K. Okuyama and K. Sakai, “Multi-boundary correlators in JT gravity,” *JHEP* **08** (2020) 126, [arXiv:2004.07555 \[hep-th\]](#).
- [36] K. Okuyama, “Eigenvalue instantons in the spectral form factor of random matrix model,” *JHEP* **03** (2019) 147, [arXiv:1812.09469 \[hep-th\]](#).
- [37] P. Saad, D. Stanford, Z. Yang, and S. Yao, “A convergent genus expansion for the plateau,” [arXiv:2210.11565 \[hep-th\]](#).
- [38] A. Blommaert, J. Kruthoff, and S. Yao, “An integrable road to a perturbative plateau,” *JHEP* **04** (2023) 048, [arXiv:2208.13795 \[hep-th\]](#).
- [39] M. L. Mehta, *Random matrices*. Elsevier, 2004.
- [40] H. Zolfi, “Firewalls from wormholes in higher genus,” [arXiv:2401.04476 \[hep-th\]](#).
- [41] L. Susskind, “The typical-state paradox: diagnosing horizons with complexity,” *Fortsch. Phys.* **64** (2016) 84–91, [arXiv:1507.02287 \[hep-th\]](#).
- [42] U. Moitra, S. K. Sake, and S. P. Trivedi, “Jackiw-Teitelboim gravity in the second order formalism,” *JHEP* **10** (2021) 204, [arXiv:2101.00596 \[hep-th\]](#).
- [43] D. L. Jafferis, D. K. Kolchmeyer, B. Mukhametzhanov, and J. Sonner, “Jackiw-Teitelboim gravity with matter, generalized eigenstate thermalization hypothesis, and random matrices,” *Phys. Rev. D* **108** no. 6, (2023) 066015, [arXiv:2209.02131 \[hep-th\]](#).
- [44] A. Belaey, F. Mariani, and T. G. Mertens, “Branes in JT (super)gravity from group theory,” [arXiv:2310.04245 \[hep-th\]](#).
- [45] D. Stanford and L. Susskind, “Complexity and Shock Wave Geometries,” *Phys. Rev. D* **90** no. 12, (2014) 126007, [arXiv:1406.2678 \[hep-th\]](#).
- [46] T. Weber, F. Haneder, K. Richter, and J. D. Urbina, “Constraining Weil-Petersson volumes by universal random matrix correlations in low-dimensional quantum gravity,” [arXiv:2208.13802 \[hep-th\]](#).
- [47] A. Blommaert, J. Kruthoff, and S. Yao, “The power of Lorentzian wormholes,” *JHEP* **10** (2023) 005, [arXiv:2302.01360 \[hep-th\]](#).
- [48] M. Usatyuk, “Comments on Lorentzian topology change in JT gravity,” [arXiv:2210.04906 \[hep-th\]](#).
- [49] L. Iliesiu, A. Levine, H. Lin, H. Maxfield, and M. Mezei, “The Non-Perturbative Hilbert Space of JT Gravity.”

- [50] A. Blommaert, T. G. Mertens, and H. Verschelde, “The Schwarzian Theory - A Wilson Line Perspective,” *JHEP* **12** (2018) 022, [arXiv:1806.07765 \[hep-th\]](#).
- [51] Z. Yang, “The Quantum Gravity Dynamics of Near Extremal Black Holes,” *JHEP* **05** (2019) 205, [arXiv:1809.08647 \[hep-th\]](#).
- [52] X. Dong, D. Marolf, P. Rath, A. Tajdini, and Z. Wang, “The spacetime geometry of fixed-area states in gravitational systems,” *JHEP* **08** (2022) 158, [arXiv:2203.04973 \[hep-th\]](#).
- [53] D. Stanford, Z. Yang, and S. Yao, “Subleading weingartens,” *JHEP* **02** (2022) 200, [arXiv:2107.10252 \[hep-th\]](#).
- [54] M. Mirzakhani, “Simple geodesics and weil-petersson volumes of moduli spaces of bordered riemann surfaces,” *Inventiones mathematicae* **167** no. 1, (2007) 179–222.
- [55] R. Dijkgraaf and E. Witten, “Developments in Topological Gravity,” *Int. J. Mod. Phys. A* **33** no. 30, (2018) 1830029, [arXiv:1804.03275 \[hep-th\]](#).
- [56] D. Stanford and E. Witten, “JT Gravity and the Ensembles of Random Matrix Theory,” [arXiv:1907.03363 \[hep-th\]](#).
- [57] K. Efetov, “Supersymmetry and theory of disordered metals,” *advances in Physics* **32** no. 1, (1983) 53–127.
- [58] A. Blommaert, T. G. Mertens, and H. Verschelde, “Eigenbranes in Jackiw-Teitelboim gravity,” [arXiv:1911.11603 \[hep-th\]](#).
- [59] L. V. Iliesiu, S. S. Pufu, H. Verlinde, and Y. Wang, “An exact quantization of Jackiw-Teitelboim gravity,” *JHEP* **11** (2019) 091, [arXiv:1905.02726 \[hep-th\]](#).
- [60] V. Chandrasekaran, R. Longo, G. Penington, and E. Witten, “An algebra of observables for de Sitter space,” *JHEP* **02** (2023) 082, [arXiv:2206.10780 \[hep-th\]](#).
- [61] S. Leutheusser and H. Liu, “Causal connectability between quantum systems and the black hole interior in holographic duality,” *Phys. Rev. D* **108** no. 8, (2023) 086019, [arXiv:2110.05497 \[hep-th\]](#).
- [62] T. G. Mertens, G. J. Turiaci, and H. L. Verlinde, “Solving the Schwarzian via the Conformal Bootstrap,” *JHEP* **08** (2017) 136, [arXiv:1705.08408 \[hep-th\]](#).
- [63] A. Blommaert, T. G. Mertens, and H. Verschelde, “Fine Structure of Jackiw-Teitelboim Quantum Gravity,” *JHEP* **09** (2019) 066, [arXiv:1812.00918 \[hep-th\]](#).
- [64] D. K. Kolchmeyer, “von Neumann algebras in JT gravity,” *JHEP* **06** (2023) 067, [arXiv:2303.04701 \[hep-th\]](#).



- [65] H. T. Lam, T. G. Mertens, G. J. Turiaci, and H. Verlinde, “Shockwave S-matrix from Schwarzian Quantum Mechanics,” *JHEP* **11** (2018) 182, [arXiv:1804.09834 \[hep-th\]](#).
- [66] D. Stanford, S. Vardhan, and S. Yao, “Scramblon loops,” [arXiv:2311.12121 \[hep-th\]](#).
- [67] Wikipedia, “Reciprocal gamma function.”.
- [68] I. Heemskerk, D. Marolf, J. Polchinski, and J. Sully, “Bulk and Transhorizon Measurements in AdS/CFT,” *JHEP* **10** (2012) 165, [arXiv:1201.3664 \[hep-th\]](#).
- [69] L. Susskind, “New Concepts for Old Black Holes,” [arXiv:1311.3335 \[hep-th\]](#).
- [70] L. Susskind, “Computational Complexity and Black Hole Horizons,” *Fortsch. Phys.* **64** (2016) 24–43, [arXiv:1403.5695 \[hep-th\]](#). [Addendum: *Fortsch.Phys.* 64, 44–48 (2016)].
- [71] D. Stanford, “More quantum noise from wormholes,” [arXiv:2008.08570 \[hep-th\]](#).
- [72] A. Blommaert, T. G. Mertens, and S. Yao, “The q-Schwarzian and Liouville gravity,” [arXiv:2312.00871 \[hep-th\]](#).
- [73] A. Blommaert, T. Mertens, and J. Papalini, “Sine dilaton gravity and double-scaled SYK.”.
- [74] D. Marolf and H. Maxfield, “Transcending the ensemble: baby universes, spacetime wormholes, and the order and disorder of black hole information,” *JHEP* **08** (2020) 044, [arXiv:2002.08950 \[hep-th\]](#).
- [75] A. Blommaert and J. Kruthoff, “Gravity without averaging,” [arXiv:2107.02178 \[hep-th\]](#).
- [76] A. Blommaert, L. V. Iliesiu, and J. Kruthoff, “Gravity factorized,” [arXiv:2111.07863 \[hep-th\]](#).
- [77] A. Blommaert, L. V. Iliesiu, and J. Kruthoff, “Alpha states demystified — towards microscopic models of AdS<sub>2</sub> holography,” *JHEP* **08** (2022) 071, [arXiv:2203.07384 \[hep-th\]](#).
- [78] P. Saad, S. H. Shenker, D. Stanford, and S. Yao, “Wormholes without averaging,” [arXiv:2103.16754 \[hep-th\]](#).
- [79] J. Kruthoff and A. Levine, “Semi-classical dilaton gravity and the very blunt defect expansion,” [arXiv:2402.10162 \[hep-th\]](#).
- [80] I. Kourkoulou and J. Maldacena, “Pure states in the SYK model and nearly-AdS<sub>2</sub> gravity,” [arXiv:1707.02325 \[hep-th\]](#).
- [81] P. Gao, D. L. Jafferis, and D. K. Kolchmeyer, “An effective matrix model for dynamical end of the world branes in Jackiw-Teitelboim gravity,” [arXiv:2104.01184 \[hep-th\]](#).

- [82] B. Post, J. van der Heijden, and E. Verlinde, “A universe field theory for JT gravity,” *JHEP* **05** (2022) 118, [arXiv:2201.08859 \[hep-th\]](#).
- [83] A. Altland and J. Sonner, “Late time physics of holographic quantum chaos,” [arXiv:2008.02271 \[hep-th\]](#).
- [84] B. Eynard, E. Garcia-Failde, P. Gregori, D. Lewanski, and R. Schiappa, “Resurgent Asymptotics of Jackiw-Teitelboim Gravity and the Nonperturbative Topological Recursion,” [arXiv:2305.16940 \[hep-th\]](#).
- [85] L. Griguolo, J. Papalini, L. Russo, and D. Seminara, “The resurgence of the plateau in supersymmetric  $N = 1$  Jackiw-Teitelboim gravity,” [arXiv:2310.06768 \[hep-th\]](#).
- [86] K. Okuyama and K. Sakai, “FZZT branes in JT gravity and topological gravity,” [arXiv:2108.03876 \[hep-th\]](#).
- [87] J. Louko and R. D. Sorkin, “Complex actions in two-dimensional topology change,” *Class. Quant. Grav.* **14** (1997) 179–204, [arXiv:gr-qc/9511023](#).