

Eigenstate Thermalization and Disorder Averaging in Gravity

Jason Pollack,^{*} Moshe Rozali,[†] James Sully,[‡] and David Wakeham,[§]

Department of Physics and Astronomy, University of British Columbia, Vancouver, British Columbia V6T 1Z1, Canada

 (Received 21 April 2020; revised 14 May 2020; accepted 9 June 2020; published 10 July 2020)

Naively, resolving the black hole information paradox requires microscopic details about quantum gravity. Recent work suggests that, instead, a unitary Page curve can be recovered by adding disorder-averaged replica instantons to the path integral, though their origin is unclear. In this Letter, we show how replica instantons and disorder averaging emerge naturally in an effective theory built from typical microscopic states. We relate replica instantons to a moment expansion of simple operators, and find a microcanonical description in terms of wormholes and Euclidean black holes.

DOI: [10.1103/PhysRevLett.125.021601](https://doi.org/10.1103/PhysRevLett.125.021601)

Introduction and summary.—Recent progress on the problem of information loss in black holes [1–3] has shed light on how semiclassical gravity calculations can be made consistent with unitarity, independent of microscopic details in the UV. (See [4–7] for related discussion.) The key step is the inclusion of *replica instantons* in the semiclassical path integral [8,9]. Replica instantons are Euclidean configurations involving correlations between several copies of the theory on distinct anti-de Sitter (AdS) boundaries. A single boundary theory with fixed couplings seems to lack such correlations, so gravitational calculations involving connected multiboundary correlators have commonly been interpreted in terms of an ensemble of theories. The goal of this Letter is to see how these correlations, and a statistical description of typical correlation functions, arise within a single unitary theory.

A statistical description of individual quantum theories and states is provided by the celebrated eigenstate thermalization hypothesis (ETH) [10–13]. The basic idea is that an isolated chaotic system appears thermal when probed only with simple (macroscopic) operators. There is still a pure state evolving unitarily in a single quantum system, but the physics is effectively coarse grained by the limited questions we can ask about the system.

Specifically, the ETH states that the matrix elements of a collection of simple operators $\{\mathcal{O}_a\}$ can be written

$$\langle E_i | \mathcal{O}_a | E_j \rangle = f_1^{(a)}(E) \delta_{ij} + e^{-S/2} f_2^{(a)}(E, \Delta E) R_{ij}^{(a)}, \quad (1)$$

for energy eigenstates $|E_i\rangle$ of a chaotic system. Here, $f_1^{(a)}$ and $f_2^{(a)}$ determine the mean and variance, respectively, of matrix elements for nearby eigenstates. In a given theory, the $R_{ij}^{(a)}$ are a fixed set of $O(1)$ numbers. If we cannot distinguish specific states, we can effectively replace the matrix elements by random variables with the correct statistics. Using this description does not mean that we are working in an ensemble of theories; instead, we are interested in the properties of *typical* states in a single theory.

In this Letter, we argue that the correct objective for the semiclassical saddle point expansion of effective field theory is reproducing the correlators of simple operators in typical microscopic states. These correlators are well described by the ETH and we can write down effective partition functions that generate their moments.

We will derive a set of Feynman rules for diagrammatically computing the partition functions for the moments of correlators. We show that, in holographic theories, the partition functions may be understood using the gravitational path integral, with higher moments related to the replica instantons of [8,9]. Assuming chaos in the microcanonical ensemble, we find a simple description of higher moments and replica instantons in terms of familiar Euclidean black holes connected by “wormholes”.

Ensembles, quantum chaos, and the ETH.—In this section, we expand on the ETH and its relationship to low-energy effective field theory. Consider a *microscopic* Hilbert space \mathcal{H} for a theory with a gravitational description. For concreteness, we work in a conformal field theory (CFT) with large central charge c . We concentrate on the subspace of states within some microcanonical energy window of width δE about energy E , denoted \mathcal{H}_E [14]. We will consider sufficiently high energies E so that the microcanonical Hilbert space has dimension $\exp[S(E, \delta E)] \sim O(e^c)$. For the remainder of this Letter, we suppress any dependence on δE .

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Funded by SCOAP³.

microcanonical basis (m, n, \dots) . The double line has an associated propagator:

$$\begin{array}{c} i \\ \bullet \\ \hline \bullet \\ n \end{array} \begin{array}{c} \longrightarrow \\ \hline \longrightarrow \\ j \\ \bullet \\ n \end{array} = \delta_{ij} \delta_{mn}.$$

States can be contracted to form a (topological) disk, with operators traced in order around the internal loop. Thus, our mean partition function is computed by

$$\begin{array}{c} i \\ \bullet \\ \text{Ⓢ} \\ \bullet \\ j \end{array} = \overline{\langle \psi_i | \mathcal{O}_a | \psi_j \rangle} = \delta_{ij} e^{-S} \text{tr}[\mathcal{O}_a].$$

The gravitational description: The partition function for a CFT at finite temperature is prepared by a path integral on $S_d \times S_1$. When a bulk dual exists, the gravitational picture is well known (see, e.g., [24–26]). At high temperatures, the leading semiclassical saddle to the gravitational partition function with boundary $S_d \times S_1$ is a Euclidean black hole. One can compute simple bulk correlation functions in this background, and find that their boundary limit matches the leading-order thermal CFT correlation function [26–28].

Note that the canonical picture is not essential to this story. With a little more effort, one can similarly find the bulk solution dual to the microcanonical partition function [29]. As long as the microcanonical width scales as $O(1) < \delta E < O(c^{1/2})$, for CFT central charge c , the projection of the bulk gravitational path integral onto a microcanonical band results in a single semiclassical bulk geometry. Thus, our Feynman diagram is identified with a black hole,

$$\begin{array}{c} i \\ \bullet \\ \text{Ⓢ} \\ \bullet \\ j \end{array} = e^{-S} \begin{array}{c} \bullet \\ \text{Ⓢ} \\ \bullet \end{array},$$

where the rhs is a diagrammatic shorthand for the microcanonical gravitational path integral computing $\text{tr}_{\mathcal{H}_E}[\mathcal{O}_a]$, and we evaluate the lhs in the same typical state, $i = j$, without summing. The partition functions are equivalent:

$$Z^{(1)}(E, J_a) = Z_{\text{Grav}}(E, J_a) \approx Z_{\text{Grav}}(\beta_E, J_a). \quad (9)$$

The first identity is holographic, and the second holds in the thermodynamic limit.

Generating functions for second moments.—So far, we have only required that the saddles of our effective field theory describe the mean microcanonical value of simple correlation functions. However, physical processes may probe slightly more fine-grained “mesoscopic” questions about the CFT, namely the higher moments. The simplest such quantities are the covariances of the distribution of $R_{ij}^{(a)}$. We can extract the covariances from products of the form

$$\langle \psi_i | \mathcal{O}_a | \psi_j \rangle \langle \psi_k | \mathcal{O}_b | \psi_l \rangle, \quad (10)$$

now involving two copies of the theory. Using Haar averages, one can check (see the Supplemental Material [30], Appendix A) that

$$\begin{aligned} & \overline{\langle \psi_i | \mathcal{O}_a | \psi_j \rangle \langle \psi_k | \mathcal{O}_b | \psi_l \rangle} \\ &= \delta_{ij} \delta_{kl} e^{-2S} \left(1 - \frac{\delta_{jk}}{e^S + 1} \right) \text{tr}[\mathcal{O}_a] \text{tr}[\mathcal{O}_b] \\ &+ \delta_{jk} \delta_{li} e^{-2S} \left(1 - \frac{\delta_{ij}}{e^S + 1} \right) \text{tr}[\mathcal{O}_a \star \mathcal{O}_b], \end{aligned} \quad (11)$$

where the traces are over \mathcal{H}_E . Likewise, the Haar averages contract operator indices only in \mathcal{H}_E . We denote this projected multiplication as $\mathcal{O}_a \star \mathcal{O}_b \equiv \mathcal{O}_a P_E \mathcal{O}_b$.

The first line in the above expression depends only on mean values of operators. At leading order, it is just the product of disconnected mean generating functions, though it receives an e^{-S} correction. The second line is a connected contribution not derivable from the mean generating function. Like the first line, it also receives an e^{-S} correction.

We can therefore write a generating function for the general second moment as

$$\begin{aligned} & Z_{ij,kl}^{(2)}(E, J_{1,a}, J_{2,b}) \\ & \equiv \sum_{a,b} J_{1,a} J_{2,b} \overline{\langle \psi_i | \mathcal{O}_a | \psi_j \rangle \langle \psi_k | \mathcal{O}_b | \psi_l \rangle} \\ &= Z_{ij}^{(1)}(E, J_{1,a}) Z_{kl}^{(1)}(E, J_{2,b}) \left(1 - \frac{\delta_{jk}}{e^S + 1} \right) \\ &+ \delta_{jk} \delta_{li} e^{-2S} Z^{(2)}(E, J_{1,a}, J_{2,b}) \left(1 - \frac{\delta_{ij}}{e^S + 1} \right), \end{aligned} \quad (12)$$

where

$$Z^{(2)}(E, J_{1,a}, J_{2,b}) \equiv \sum_{a,b} J_{1,a} J_{2,b} \text{tr}[\mathcal{O}_a \star \mathcal{O}_b]. \quad (13)$$

Feynman rules for the second moment partition function: With the Feynman rules we introduced to compute the mean correlator, we can already compute the leading order contribution on each line of (11):

$$\begin{aligned} & \begin{array}{c} i \\ \bullet \\ \text{Ⓢ} \\ \bullet \\ j \end{array} \begin{array}{c} \bullet \\ \text{Ⓢ} \\ \bullet \\ k \end{array} \begin{array}{c} l \\ \bullet \\ \text{Ⓢ} \\ \bullet \\ k \end{array} = \delta_{ij} \delta_{kl} e^{-2S} \text{tr}[\mathcal{O}_a] \text{tr}[\mathcal{O}_b] \\ & \begin{array}{c} i \\ \bullet \\ \text{Ⓢ} \\ \bullet \\ j \end{array} \begin{array}{c} \bullet \\ \text{Ⓢ} \\ \bullet \\ k \end{array} = \delta_{jk} \delta_{li} e^{-S} \text{tr}[\mathcal{O}_a \star \mathcal{O}_b]. \end{aligned}$$

To compute the subleading terms, we need to introduce a new vertex which resums to generate the factor of $1/(e^S + 1)$ and enforces that the state index passing along the lines it joins are equal. This is achieved by

$$\begin{array}{c} i \\ \curvearrowright \\ j \end{array} \begin{array}{c} \curvearrowleft \\ l \\ k \end{array} = -e^{-S} \delta_{ij} \delta_{kl} \delta_{li},$$

whose resummation gives

$$\sum_n \begin{array}{c} i \\ \curvearrowright \\ j \end{array} \begin{array}{c} \curvearrowleft \\ l \\ k \end{array} = \begin{array}{c} i \\ \curvearrowright \\ j \end{array} \begin{array}{c} \curvearrowleft \\ l \\ k \end{array} = -\frac{\delta_{ij} \delta_{kl} \delta_{li}}{e^S + 1}.$$

We thus have Feynman diagrams for the corrections:

$$\begin{array}{c} i \\ \curvearrowright \\ j \end{array} \begin{array}{c} \curvearrowleft \\ l \\ k \end{array} \begin{array}{c} i \\ \curvearrowright \\ j \end{array} \begin{array}{c} \curvearrowleft \\ l \\ k \end{array}, \quad \begin{array}{c} i \\ \curvearrowright \\ j \end{array} \begin{array}{c} \curvearrowleft \\ l \\ k \end{array} \begin{array}{c} i \\ \curvearrowright \\ j \end{array} \begin{array}{c} \curvearrowleft \\ l \\ k \end{array}.$$

The gravitational description: The gravitational description of $Z^{(2)}$ requires just a slight elaboration on the case of $Z^{(1)}$. As before, where the microcanonical partition function was given at leading order by the microcanonical black hole saddle, the insertion of another simple operator in the trace does not shift to another saddle and we can compute the correlator using the same bulk solution. Since the same energy runs between both operators, they must be equally spaced on opposite sides of the circular Euclidean-time boundary. For further discussion of the placement of operators on the thermal circle, see the Supplemental Material [30], Appendix B.

In the canonical language, we have the approximate identity (in the thermodynamic limit)

$$\text{tr}_{\mathcal{H}_E} [\mathcal{O}_a \star \mathcal{O}_b] \approx \text{tr}_{\mathcal{H}} [e^{-\beta_E H/2} \mathcal{O}_a e^{-\beta_E H/2} \mathcal{O}_b]. \quad (14)$$

The symmetric insertion of Euclidean evolution between each operator approximately projects them all onto the same microcanonical window, \mathcal{H}_E , where the sum over energies localizes. Thus, we can write

$$\begin{aligned} Z^{(2)}(E, J_{1,a}, J_{2,b}) &= Z_{\text{Grav}}(E, J_{1,a}, J_{2,b}) \\ &\approx Z_{\text{Grav}}(\beta_E, J_a, J_b^{(1/2)}), \end{aligned} \quad (15)$$

where $J_b^{(1/2)}$ is the source for the operator $\mathcal{O}_b(\beta/2) = e^{-\beta_E H/2} \mathcal{O}_b e^{\beta_E H/2}$. Our immediate takeaway is that the Euclidean wormhole needed to compute $Z^{(2)}$ is just the standard wormhole for the microcanonical (or thermal) black hole.

Furthermore, while we have no direct gravitational interpretation of the corrections to the leading terms, we have suggested that they might be thought of as topologically nontrivial ‘‘wormholes’’ that glue the geometries together, such as

$$\begin{array}{c} \text{---} \text{---} \\ \text{---} \text{---} \\ \text{---} \text{---} \\ \text{---} \text{---} \end{array} \begin{array}{c} \text{---} \text{---} \\ \text{---} \text{---} \\ \text{---} \text{---} \\ \text{---} \text{---} \end{array} = e^{-2S} \begin{array}{c} \text{---} \text{---} \\ \text{---} \text{---} \\ \text{---} \text{---} \\ \text{---} \text{---} \end{array}.$$

See [8] for a related discussion of ‘‘handles’’ joining replica instantons. In Appendix C [30], we work through examples of the squared partition function and second Rényi entropy.

Higher moments.—We can similarly compute higher moments of our correlation functions, with details relegated to Appendix A [30]. In short, to calculate

$$\overline{\prod_{m=1}^n \langle \psi_{i_m} | \mathcal{O}_{a_m} | \psi_{j_m} \rangle} \quad (16)$$

at leading order, we sum over Feynman diagrams showing how identical states are joined to form traces of the operators. For example, the fourth moment has a term

$$\begin{array}{c} j_1 \quad i_1 \\ \text{---} \text{---} \\ \text{---} \text{---} \\ \text{---} \text{---} \\ i_3 \quad j_3 \end{array} \begin{array}{c} i_2 \quad j_2 \\ \text{---} \text{---} \\ \text{---} \text{---} \\ \text{---} \text{---} \\ i_4 \quad j_4 \end{array}.$$

To compute the correct normalization, we must resum an infinite series of contributions from vertices linking the boundaries. We showed above how to compute the leading order corrections in terms of vertices that join two boundaries together. Further corrections take the form of k -point vertices (see the Supplemental Material [30], Appendix A 1), with

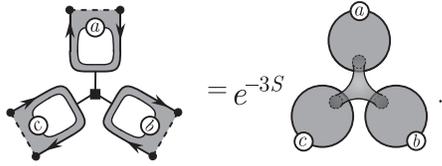
$$\begin{array}{c} i_{m_1} \\ \text{---} \text{---} \\ \text{---} \text{---} \\ \text{---} \text{---} \\ i_{m_k} \end{array} = g_k(S) \delta_{i_{m_1} \dots i_{m_k}}^{(k)}.$$

The gravitational description: Each leading-order term needed to compute the higher moments is just a product of microcanonical traces, and will be calculated by the same gravitational saddles as the microcanonical black hole of energy E without operator insertions. For each trace with p operators inserted, we space them equally in order of the trace around the Euclidean time circle. The canonical picture is likewise simple. The canonical generating function for the trace with p operators is just given by

$$Z_{\text{Grav}}(\beta_E, \{J_{a_m}^{(m/p)}\}_{m=0}^{p-1}), \quad (17)$$

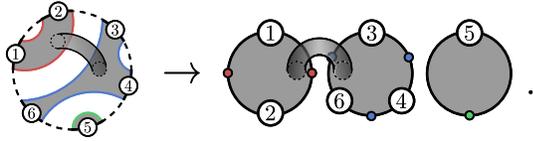
where $J_{a_m}^{(m/p)}$ sources $e^{-\beta_E H m/p} \mathcal{O}_{a_m} e^{\beta_E H m/p}$.

As in the case of the second moment, we can view the subleading corrections to each trace as k -boundary wormholes joining the true gravitational geometries together, for instance



Again, we have no true gravitational solution dual to these corrections, but regard it as a heuristic description to motivate further work.

Discussion.—In [8], statistical averages for traces of operators were computed using wormholes in a genuine ensemble of JT gravity theories. We instead perform these traces with respect to (microcanonically typical) pure states of a single chaotic theory. As a result, our connected replica wormholes are just standard microcanonical wormholes joined by “handles.” Our geometries are related to [8] by gluing distinct copies seamlessly together via bulk branes:



Other ensembles: We use a different ensemble from [8], and hence arrive at different bulk geometries. We have argued that, for a microcanonical window in a chaotic system, the CUE is the right statistical ensemble for a short-time observer. Adopting a different notion of typicality will give corrections to the moments of our distribution, and require a modified gravitational interpretation (see [31–33] and references therein).

Moreover, by restricting ourselves to states within a narrow microcanonical band, we have avoided questions about the dependence of the moments on energy differences. Understanding the relevant bulk geometries for the more general case is of obvious interest.

RG and EFT: Our notion of coarse-graining is different from the standard Wilsonian RG perspective, where UV degrees of freedom are “integrated out.” It would be illuminating to directly relate the integrating out of microscopic splittings at high energy to other approaches to renormalization and coarse graining [8,34–40].

Finally, we have employed a skeletal notion of EFT, concentrating on semiclassical bulk saddles rather than the full gravitational field theory. Relatedly, we have focused on short-time physics, when states in the Schrödinger picture have not had time to explore atypical corners of Hilbert space. Extending our framework in either direction may reveal connections to previous work on ETH and gravity at late times [23,26,41–43].

We thank Mark Van Raamsdonk for useful discussions. We are supported by NSERC, UBC, the Simons Foundation, and by readers like you.

*jpollack@phas.ubc.ca

†rozali@phas.ubc.ca

‡sully@phas.ubc.ca

§daw@phas.ubc.ca

- [1] G. Penington, [arXiv:1905.08255](#).
- [2] A. Almheiri, N. Engelhardt, D. Marolf, and H. Maxfield, *J. High Energy Phys.* **12** (2019) 063.
- [3] A. Almheiri, R. Mahajan, J. Maldacena, and Y. Zhao, *J. High Energy Phys.* **03** (2020) 149.
- [4] C. Akers, N. Engelhardt, and D. Harlow, [arXiv:1910.00972](#).
- [5] M. Rozali, J. Sully, M. Van Raamsdonk, C. Waddell, and D. Wakeham, *J. High Energy Phys.* **05** (2020) 004.
- [6] A. Almheiri, R. Mahajan, and J. Maldacena, [arXiv:1910.11077](#).
- [7] A. Almheiri, R. Mahajan, and J.E. Santos, [arXiv:1911.09666](#).
- [8] G. Penington, S.H. Shenker, D. Stanford, and Z. Yang, [arXiv:1911.11977](#).
- [9] A. Almheiri, T. Hartman, J. Maldacena, E. Shaghoulian, and A. Tajdini, *J. High Energy Phys.* **05** (2020) 013.
- [10] J.M. Deutsch, *Phys. Rev. A* **43**, 2046 (1991).
- [11] M. Srednicki, *Phys. Rev. E* **50**, 888 (1994).
- [12] M. Srednicki, *Ann. N.Y. Acad. Sci.* **755**, 757 (1995).
- [13] M.A. Srednicki, *J. Phys. A* **32**, 1163 (1999).
- [14] Here we are considering the CFT quantized on a sphere of radius R , where energy is related to conformal dimension Δ by $E \sim R\Delta$.
- [15] In the Schrödinger picture, an initial state which is typical in the microcanonical window may become atypical after exponentially long times due to quantum ergodicity.
- [16] Crudely, a typical energy splitting is order $\delta E/e^S$, which requires time $\delta t \sim \hbar e^S/\delta E$ to probe.
- [17] That is, we are working in the circular unitary ensemble (CUE).
- [18] Note that while entries for an individual matrix are iid, matrices for different operators will typically have a smooth covariance, $\overline{R_{ij}^{(a)} R_{kl}^{(b)}} \equiv \delta_{il} \delta_{jk} \sigma_2^{(ab)}$.
- [19] L. D’Alessio, Y. Kafri, A. Polkovnikov, and M. Rigol, *Adv. Phys.* **65**, 239 (2016).
- [20] The operators $\{\mathcal{O}_a\}$ do not form an algebra, since low-energy observers cannot measure arbitrary products. Our generating functions give expectations in the linear span of $\{\mathcal{O}_a\}$ only.
- [21] P. Brouwer and C. Beenakker, *J. Math. Phys. (N.Y.)* **37**, 4904 (1996).
- [22] There are parallels between our calculations and diagrammatics and those of Ref. [23]. Our results apply to typical microcanonical states in an arbitrary chaotic theory, while Ref. [23] considers a genuine ensemble of Hamiltonians in JT gravity.
- [23] P. Saad, [arXiv:1910.10311](#).
- [24] E. Witten, *Adv. Theor. Math. Phys.* **2**, 505 (1998).
- [25] G. T. Horowitz and V.E. Hubeny, *Phys. Rev. D* **62**, 024027 (2000).
- [26] J.M. Maldacena, *J. High Energy Phys.* **04** (2003) 021.
- [27] E. Keski-Vakkuri, *Phys. Rev. D* **59**, 104001 (1999).
- [28] D.T. Son and A.O. Starinets, *J. High Energy Phys.* **09** (2002) 042.
- [29] D. Marolf, *J. High Energy Phys.* **09** (2018) 114.

- [30] See the Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevLett.125.021601> for further details on Haar averaging, the relation between canonical and microcanonical ensembles, and some example calculations.
- [31] V. Balasubramanian, M. Berkooz, S. F. Ross, and J. Simon, *Classical Quantum Gravity* **31**, 185009 (2014).
- [32] P. Saad, S. H. Shenker, and D. Stanford, [arXiv:1806.06840](https://arxiv.org/abs/1806.06840).
- [33] J. S. Cotler, G. Gur-Ari, M. Hanada, J. Polchinski, P. Saad, S. H. Shenker, D. Stanford, A. Streicher, and M. Tezuka, *J. High Energy Phys.* **05** (2017) 118; **09** (2018) 2(E).
- [34] E. A. Calzetta and B.-L. B. Hu, *Nonequilibrium Quantum Field Theory*, Cambridge Monographs on Mathematical Physics (Cambridge University Press, Cambridge, England, 2008).
- [35] N. Bao, A. Chatwin-Davies, J. Pollack, and G. N. Remmen, [arXiv:1911.10207](https://arxiv.org/abs/1911.10207).
- [36] N. Lashkari, *J. High Energy Phys.* **01** (2019) 059.
- [37] O. Kabernik, J. Pollack, and A. Singh, *Phys. Rev. A* **101**, 032303 (2020).
- [38] R. Blume-Kohout, H. K. Ng, D. Poulin, and L. Viola, *Phys. Rev. Lett.* **100**, 030501 (2008).
- [39] A. Almheiri, X. Dong, and D. Harlow, *J. High Energy Phys.* **04** (2015) 163.
- [40] Y. Nomura, [arXiv:1911.13120](https://arxiv.org/abs/1911.13120).
- [41] J. L. F. Barbon and E. Rabinovici, *J. High Energy Phys.* **11** (2003) 047.
- [42] J. L. F. Barbon and E. Rabinovici, *Fortschr. Phys.* **52**, 642 (2004).
- [43] J. L. F. Barbon and E. Rabinovici, *Fortschr. Phys.* **62**, 626 (2014).