Title: Free Energy from Replica Wormholes

Speakers: Netta Engelhardt

Collection: Quantum Gravity 2020

Date: July 16, 2020 - 11:15 AM

URL: http://pirsa.org/20070014

Free Energy from Replica Wormholes

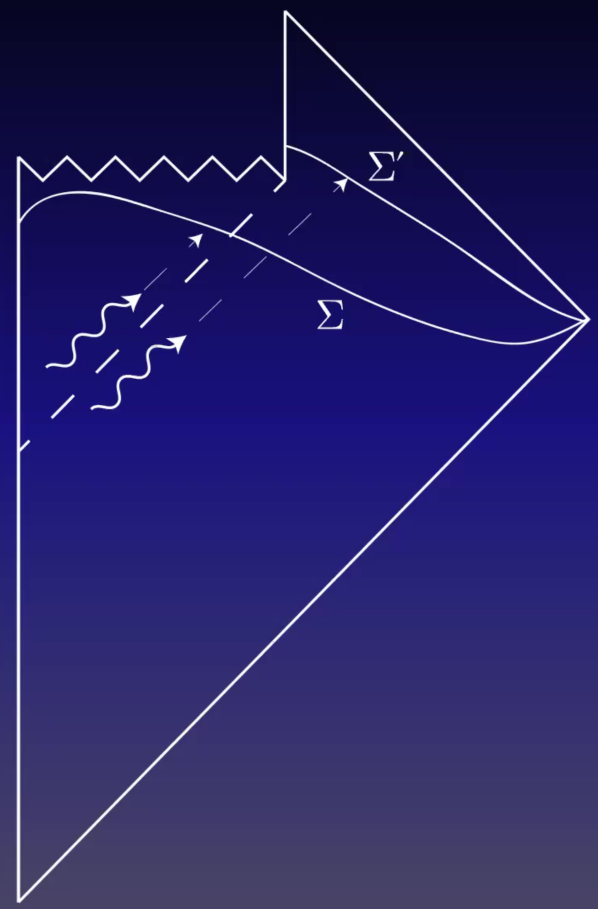
Netta Engelhardt

MIT

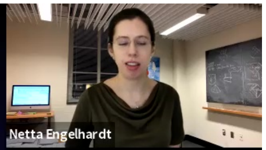
QG 2020

Some Recent New Insights

- The past year has seen a lot of progress on the black hole information paradox
- Insights from these developments have started to teach us more about gravity in general
- In particular, the inclusion of replica wormholes in the replica trick has shed new light on old insights about Euclidean wormholes and ensembles in gravity Coleman; Giddings, Strominger; Maldacena, Maoz; on the modern side, Marolf Maxfield; Giddings Turiaci,...
- To set the stage, I'll review some of these developments.



Some Recent New Insights



- The past year has seen a lot of progress on the black hole information paradox
- Insights from these developments have started to teach us more about gravity in general
- In particular, the inclusion of replica wormholes in the replica trick has shed new light on old insights about Euclidean wormholes and ensembles in gravity Coleman; Giddings, Strominger; Maldacena, Maoz; on the modern side, Marolf Maxfield; Giddings Turiaci,...
- To set the stage, I'll review some of these developments. Then I'll discuss some new work on replica wormholes with S. Fischetti and A. Maloney (on the arXiv last night)

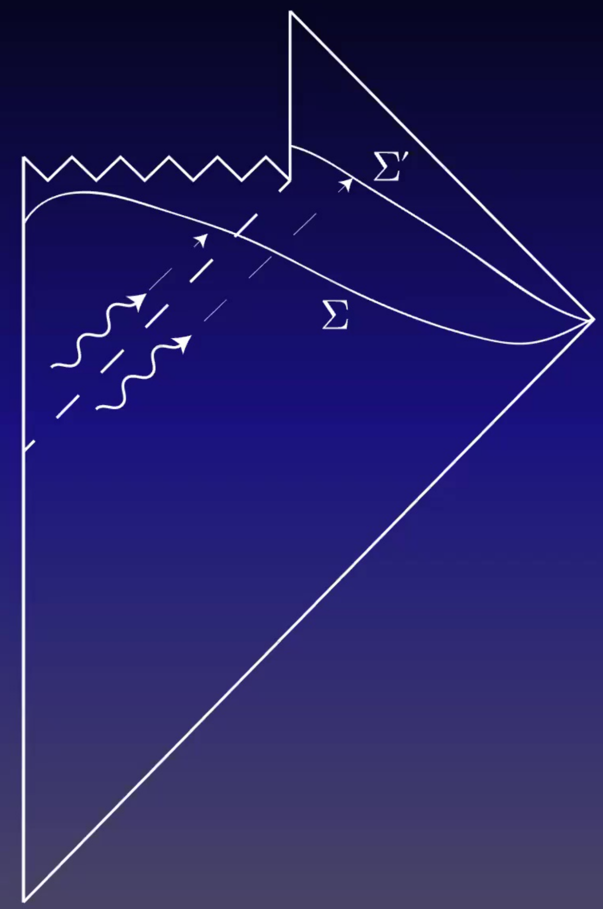
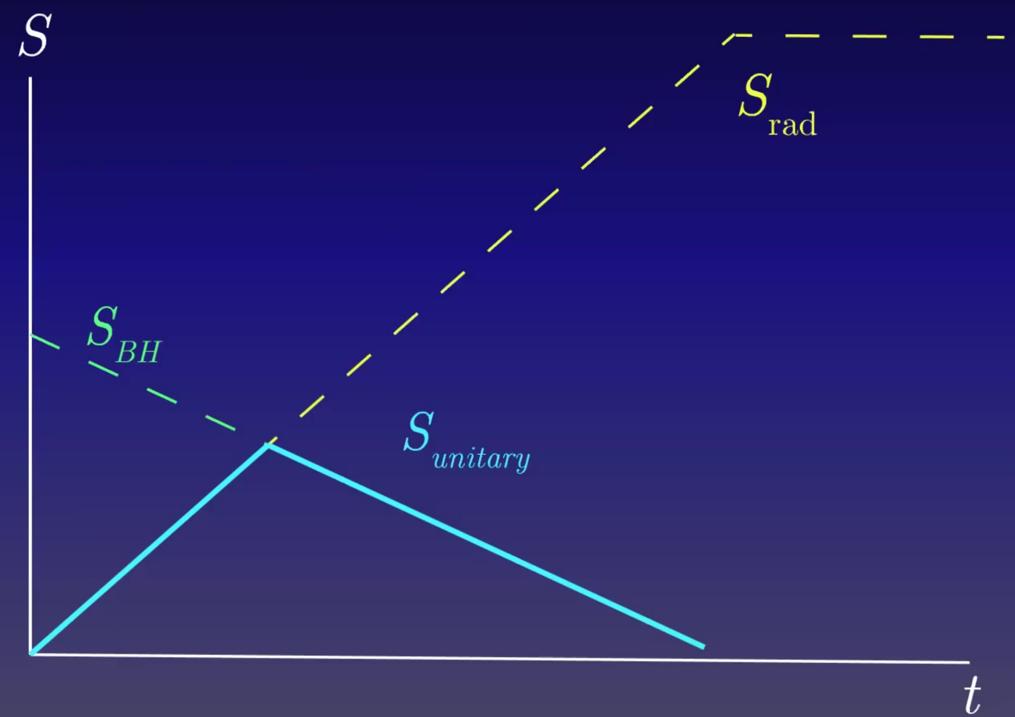
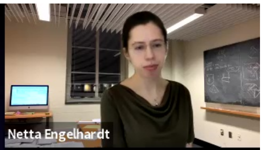


Table of Contents

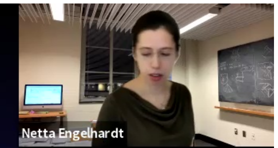
New Developments: Goal

The new developments were initially Penington; Almheiri, NE, Marolf, Maxfield centered around one main goal: to derive the Page curve for an evaporating black hole.





New Developments: Insight

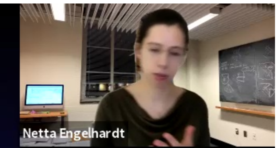


This has been a goal in quantum gravity for several decades, so what's new?

1. **Not** using Hawking's formula for computing the entropy. Instead using a holographic formula.

2. Forcing a large AdS black hole to evaporate.

Entropy Formula



Instead of using

$$S = -tr\rho \ln \rho$$

we'll use

$$S = \min \operatorname{ext} \left[\frac{\text{Area}}{4G\hbar} + S_{\text{out}} \right]$$

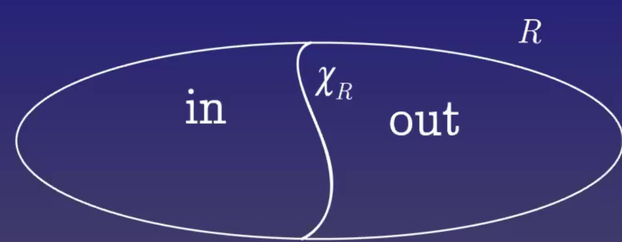
To be a little more explicit

The QES prescription NE, Wall:

$$S_{vN}[\rho_R] = \frac{\text{Area}[\chi_R]}{4G\hbar} + S_{\text{out}}[\chi_R] = S_{\text{gen}}[\chi_R]$$

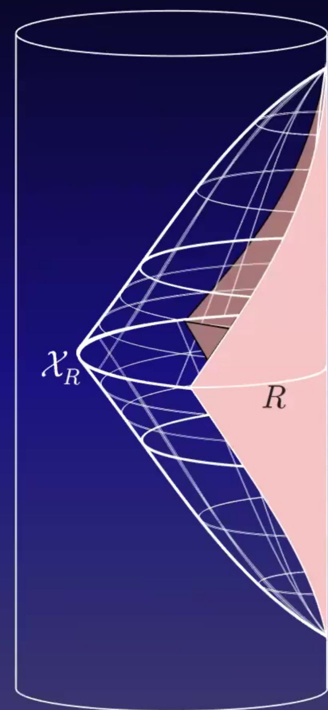
where χ_R is the minimal-S_{gen} surface that extremizes S_{gen}. (In the classical case we extremize just the area RT, HRT)

For example, for a higher-dimensional system, the quantum extremal surface (QES) of a region R:



The Entanglement Wedge

The region between χ_R and R is the *entanglement wedge*:

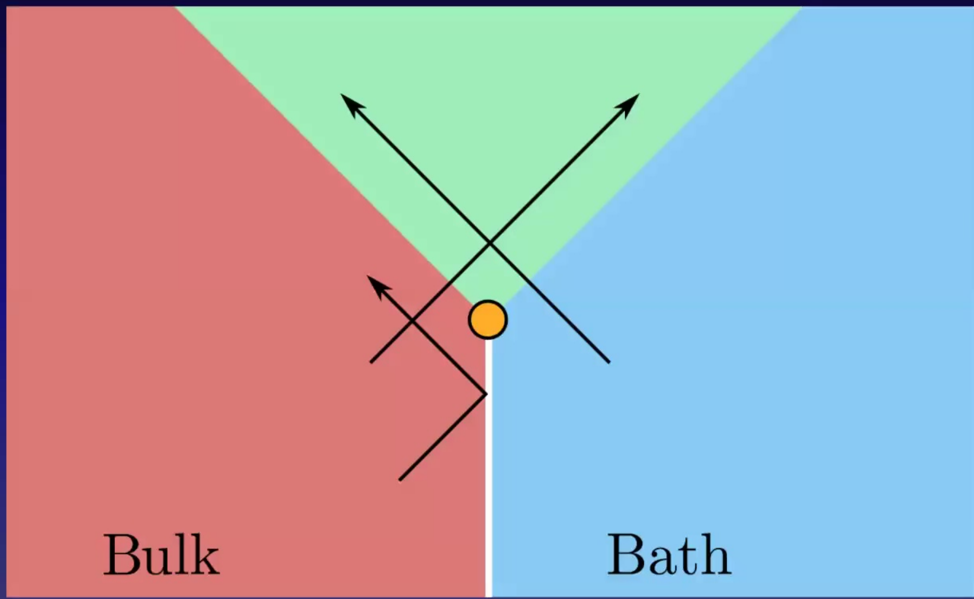


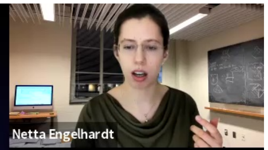
Subregion duality is the statement that the entanglement wedge of R is dual to R the really correct statement involves the Hilbert space and operator algebra on $D[R]$;

Wall; van Raamsdonk; Dong, Harlow, Wall,...

Netta Engelhardt

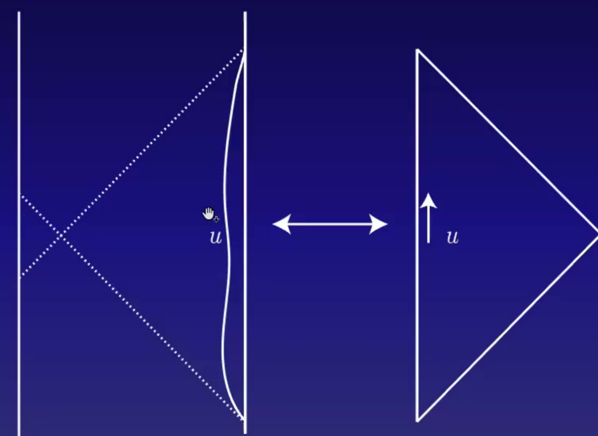
Evaporating an AdS Black Hole





Evaporating the Black Hole

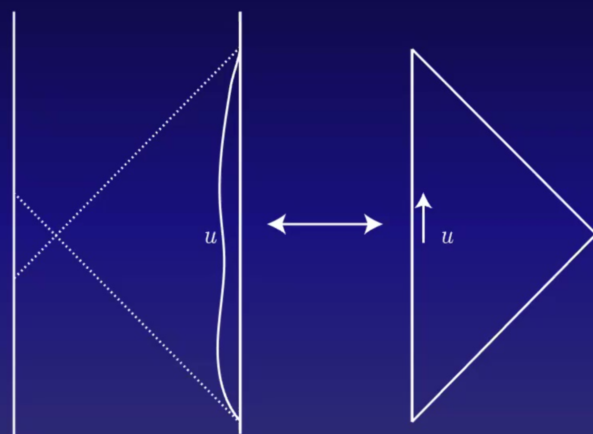
We work in JT gravity. To evaporate the black hole, we consider an auxiliary (B)CFT in flat space at zero temperature.



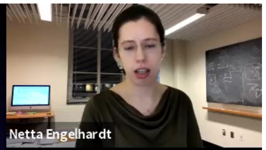
We couple the two systems (a quantum quench) at physical time u and then evolve them forwards in time. This results in a shockwave propagating into the bulk.

Evaporating the Black Hole

We work in JT gravity. To evaporate the black hole, we consider an auxiliary (B)CFT in flat space at zero temperature.

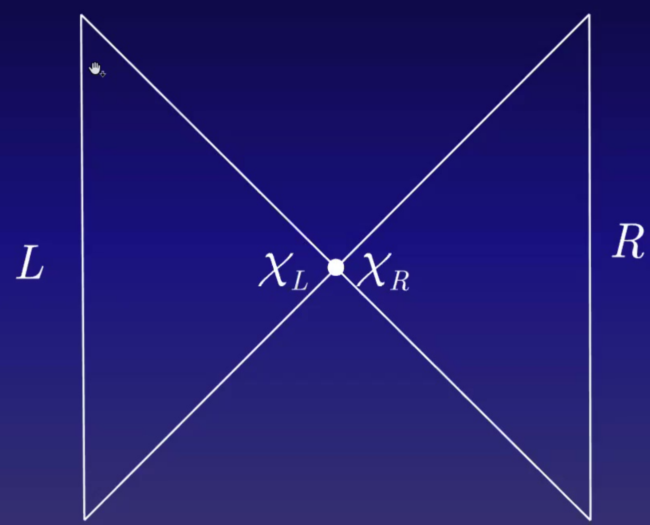


We couple the two systems (a quantum quench) at physical time u and then evolve them forwards in time. This results in a shockwave propagating into the bulk. Now we are ready to compute entanglement entropies.



Main Takeaway

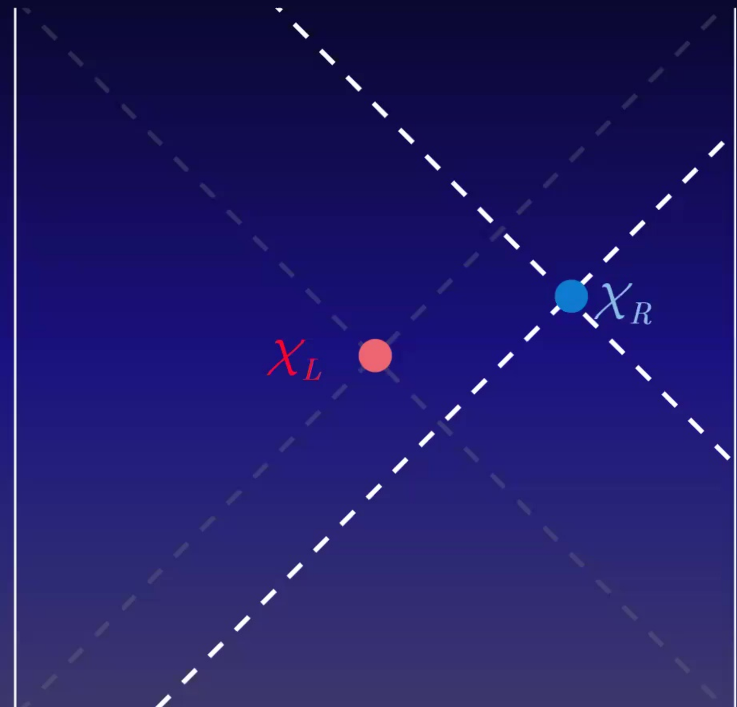
Before evaporating the black hole, in the 2-sided case the left and right QESs are just the bifurcation surface:



As we evaporate the black hole, initially χ_R moves continuously in a spacelike direction.

Main Takewaway

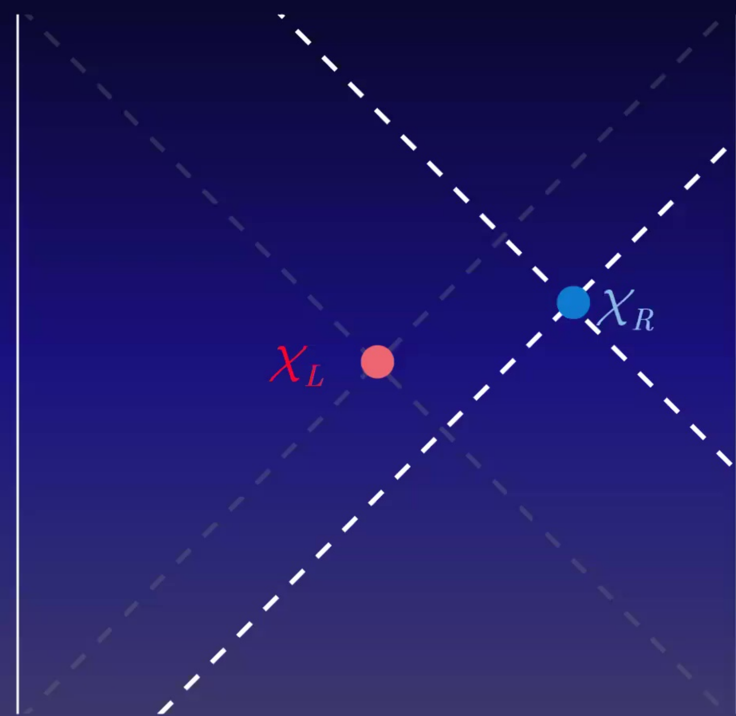
At late times, a branch of QESs with *no classical counterpart* begins to dominate:



After the transition, the dominant right QES is far from the left QES. The effect of the transition is a unitary Page curve in the bulk.

Main Takeaway

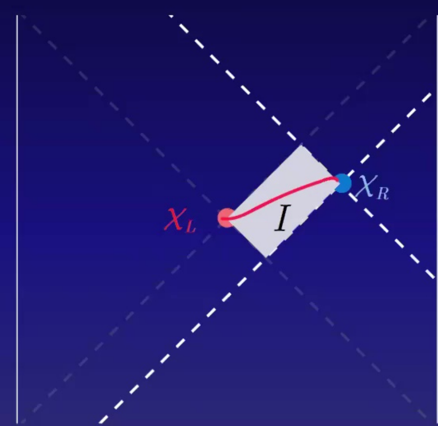
At late times, a branch of QESs with *no classical counterpart* begins to dominate:



After the transition, the dominant right QES is far from the left QES. The effect of the transition is a unitary Page curve in the bulk.
Note that if we throw the radiation into the left side, χ_L will move to χ_R.

The Quantum Extremal Island

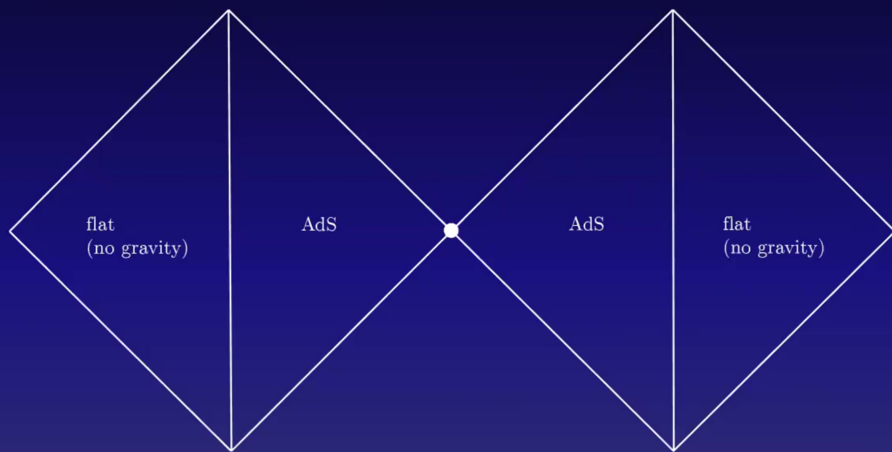
Hypothesis in Hayden, Penington, Almheiri, Mahajan, Maldacena, Zhao: the region between χ_L and χ_R corresponds to the radiation. In a sense that I will not make precise here, it is the entanglement wedge of the radiation.



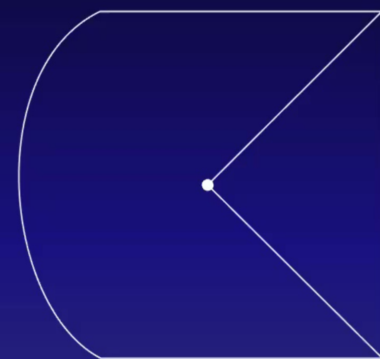
"Quantum Extremal Island" formula: Almheiri, Mahajan, Maldacena, Zhao

$$S[\rho_R] = \min \text{ext} \left[\frac{\text{Area}[\partial I]}{4G} + S[I \cup R] \right]$$

Subsequent Developments: Toy Models



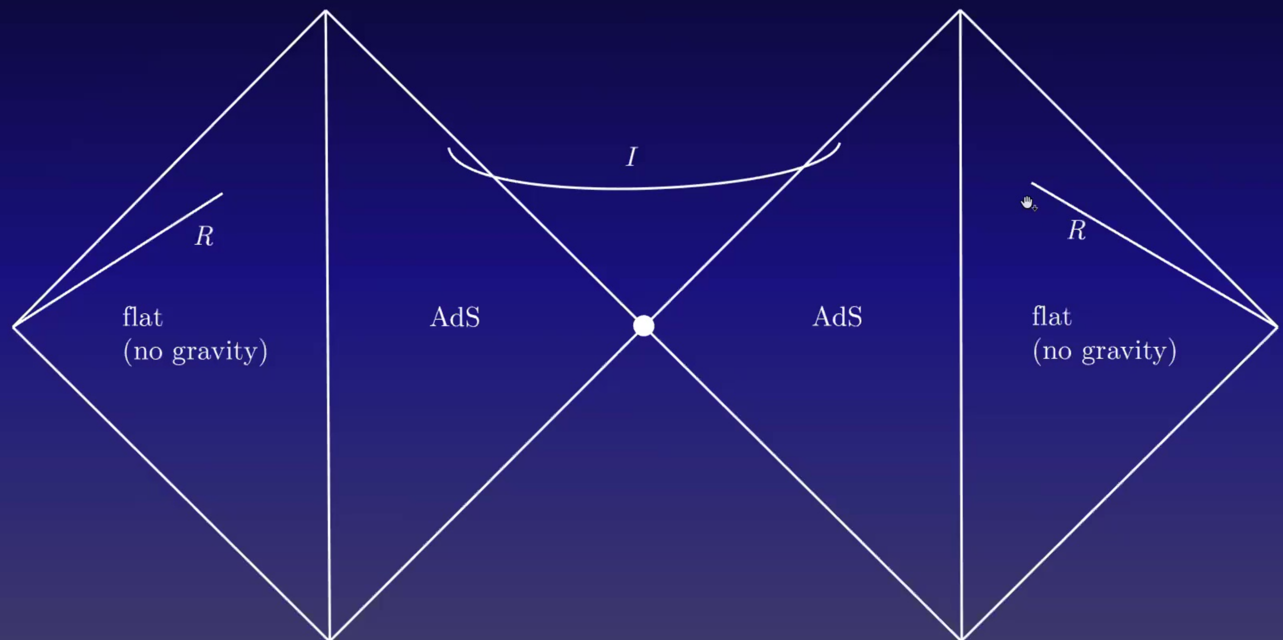
Almheiri et al "East Coast model"



Shenker et al "West Coast model"

Subsequent Developments: Deriving the QEI

The entropy of a subset of the radiation means R is part of the non-gravitational spacetime region:



Almheiri et al

Review: the S_{vN} Replica Trick

Forget about gravity for a moment.

The nth Renyi entropy of some ρ_R is defined:

$$S_n[\rho_R] = \frac{1}{1-n} \ln \mathrm{Tr} \rho_R^n$$

and S_{vN} can be computed by taking the $n \to 1$ limit.

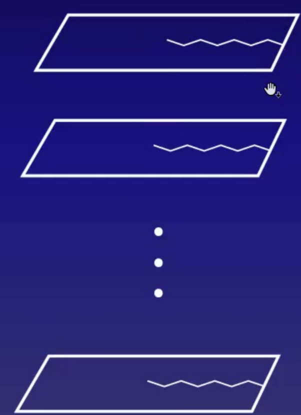
Working in Euclidean signature now, we can write the state in terms of the Euclidean partition function, which yields:

$$S_{\mathrm{vN}}[\rho_R] = \lim_{n \to 1} \frac{1}{1-n} \left(\ln Z(B_n) - n \ln Z(B) \right)$$

where $Z(B)$ is the partition function on the space B and $Z(B_n)$ is a partition function on B_n, an n-sheeted geometry consisting of n copies of B cut along R and cyclically identified along R.

Review: the S_{vN} Replica Trick

$$S_{\rm vN}[\rho_R] = \lim_{n\to 1} \frac{1}{1-n}\left(\ln Z(B_n) - n \ln Z(B)\right)$$



Review: S_{vN} Trick from the GPI

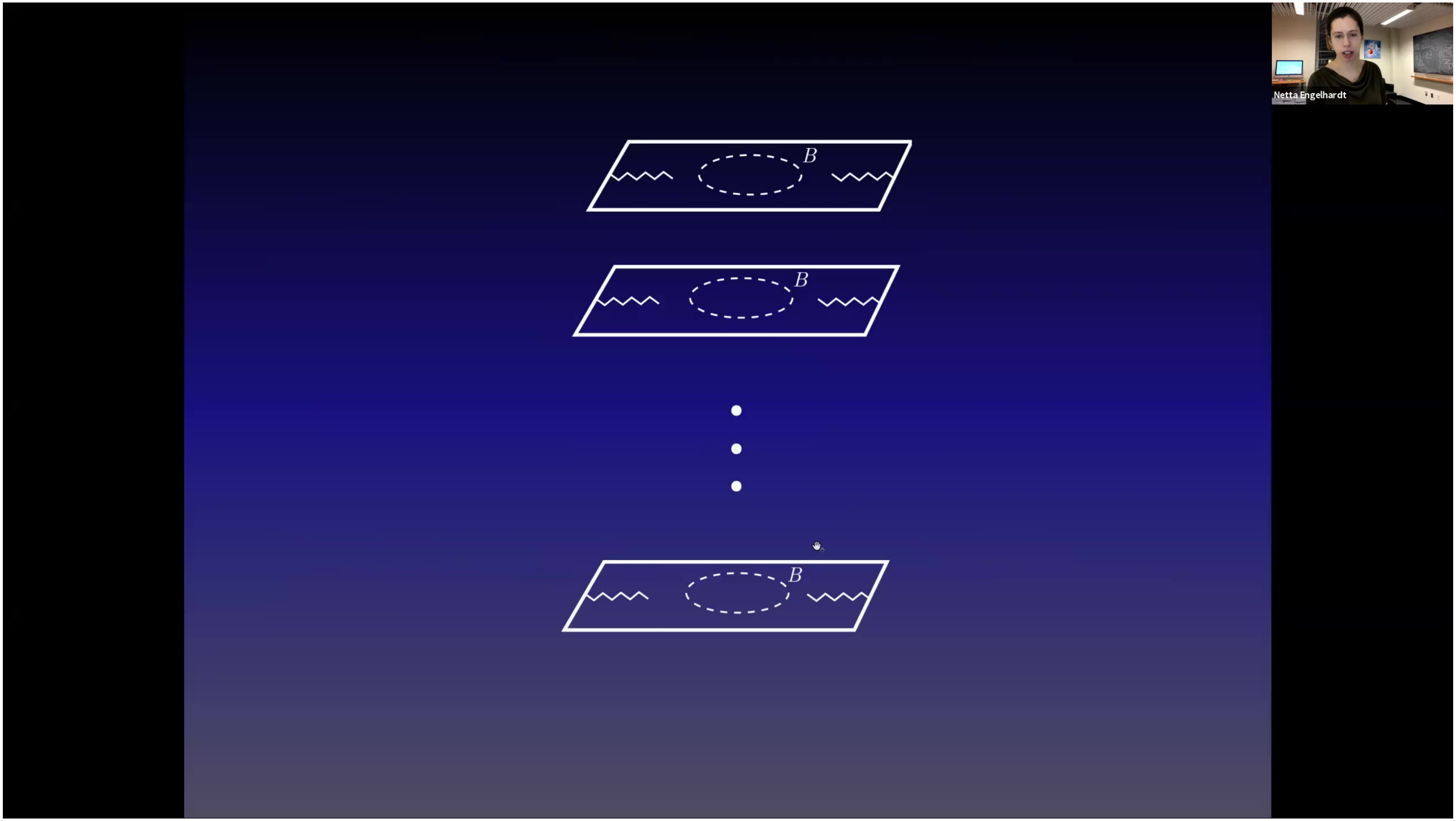
Getting back to gravity, the idea is (again, sticking to the Euclidean case) Lewkowycz-Maldacena: replace Z by the gravitational path integral:

$$Z(B) \to \mathcal{P}(B) = \int_{\partial M = B} Dg e^{-S}$$

and

$$Z(B_n) \to \mathcal{P}(B_n) = \int_{\partial M = B_n} Dg e^{-S}$$

In e.g. the East Coast model, the identified region R is not in the gravitational region, so the GPI is over n copies of the AdS boundary (not identified).

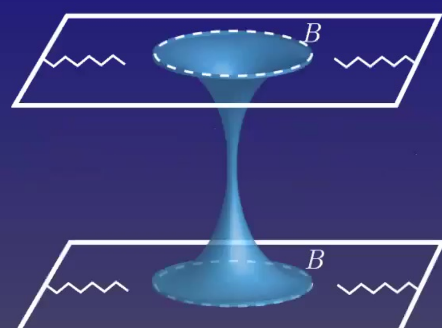

B
B
B
Netta Engelhardt

Subsequent Developments: Deriving the QEI

So the GPI $\mathcal{P}(B^n)$ is on n copies of the boundary (glossing over details involving boundary conditions). In general we have

$$\mathcal{P}(B^n) = \mathcal{P}(B)^n + \sum_{\substack{\text{connected}\\ \text{topologies}}} ,$$

Whether connected topologies should be included or not was discussed at length in the past Coleman; Giddings, Strominger; Maldacena, Maoz.
It was found by Almheiri et al; see also Shenker et al for the other model that after the Page time, the contribution from connected topologies dominates.



Subsequent Developments: Deriving the QEI

Should those topologies be included?

Those topologies are precisely what gives rise to the QEI formula. We'd like to include them, because that gives us a unitary answer. But including them means that

$$\mathcal{P}(B^n) \neq \mathcal{P}(B)^n$$

and subsequently, if we take $\mathcal{P}(B) = Z(B)$,

$$Z(B^n) \neq Z(B)^n.$$

One possibility that has long been discussed in the literature is that the GPI computes an ensemble average:

$$\mathcal{P}(B) = \overline{Z(B)}$$

Aside: averaged $Z(B)$

Normally what we mean by ensemble or disorder averaging is that the partition function is $Z_g(B)$ is defined wrt some choice of coupling constants g sampled from a distribution $P(g)$, and

$$\overline{Z(B)} = \int dg P(g) Z_g(B)$$

If this is what the GPI is computing in gravity, though, it's not obvious what an individual $Z_g(B)$ is. We'll be mostly agnostic about whether the GPI is ensemble averaging or not and focus instead on whether connected topologies contribute or not

Lessons

- If we want to take these replica wormholes seriously (which we do), then we need to understand better what kind of calculation the path integral is doing

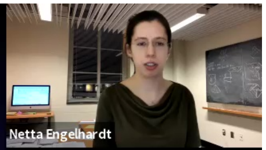
- If it's averaging, what is it averaging over?

- What are the implications of this on other observables? Can we see signatures of replica wormholes beyond the Renyi entropies?

- For instance, we might ask about the generating functional $\ln Z$ (or equivalently the free energy $-T \ln Z$.

Table of Contents

Computing $\ln Z$



If $\mathcal{P}(B)$ receives contributions from connected topologies, then it is in principle possible that there is a big difference between

$$\ln \overline{Z} = \ln \mathcal{P}(B) \qquad \text{vs} \qquad \overline{\ln Z}$$

The first ("annealed") computes an average over the random variables defining a particular instance of the ensemble, which suggests an interpretation as the partition function of a theory in which the random variables are allowed to equilibriate.

The second ("quenched") computes an average over the free energies of constituents of the ensemble, i.e. where the random variables are not allowed to equilibrate.

Computing $\ln Z$

If $\mathcal{P}(B)$ receives contributions from connected topologies, then it is in principle possible that there is a big difference between

$$\ln \overline{Z} = \ln \mathcal{P}(B) \qquad \text{vs} \qquad \overline{\ln Z}$$

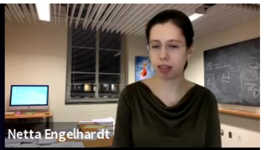
The first ("annealed") computes an average over the random variables defining a particular instance of the ensemble, which suggests an interpretation as the partition function of a theory in which the random variables are allowed to equilibriate.

The second ("quenched") computes an average over the free energies of constituents of the ensemble, i.e. where the random variables are not allowed to equilibrate. This one is the one we care about.

If $\overline{Z(B^n)} \neq \overline{Z(B)}^n$, we might also expect that $\overline{\ln Z(B)} \neq \ln \overline{Z(B)}$.

But how on earth do we make sense of $\overline{\ln Z}$ from the GPI?

The $\ln Z$ Replica Trick



We can compute $\ln Z$ via a replica trick (the condensed matter theorists have been doing this for decades)

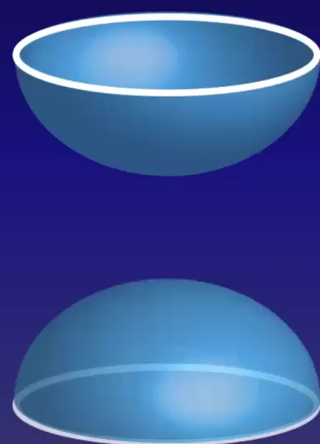
$$\ln Z(B) = \lim_{m \to 0} \frac{1}{m} \left(Z(B^m) - 1 \right)$$

This we can compute from the GPI using $\mathcal{P}(B^m)$:

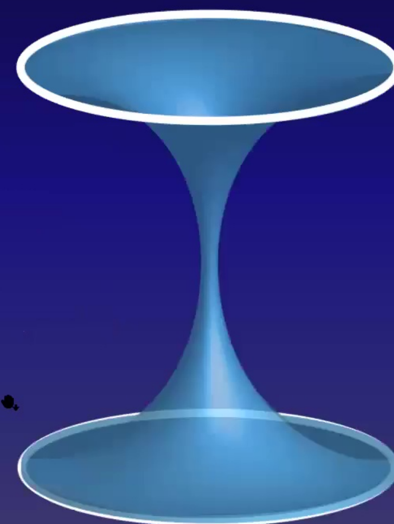
$$\overline{\ln Z(B)} = \lim_{m \to 0} \frac{1}{m} \left(\mathcal{P}(B^m) - 1 \right)$$

The $\overline{\ln Z}$ Replica Trick

If replica wormholes can contribute nontrivially in the $m \to 0$ limit, then this is very different from the annealed free energy $\ln \mathcal{P}(B)$, which would correspond to only including disconnected topologies.



Annealed: $\ln \mathcal{P}(B)$
only includes disconnected topology.

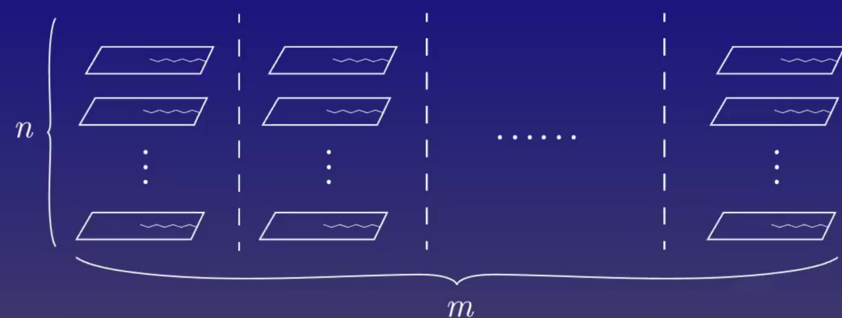


Quenched: $\lim_{m \to 0} \frac{1}{m} \left(\mathcal{P}(B^m) - 1\right)$;
can include connected topologies.

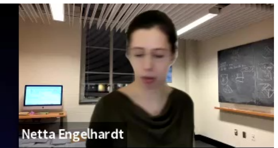
The $\overline{\ln Z}$ Replica Trick

So, for example, to correctly compute the von Neumann entropy in a holographic theory, we need to do two replica tricks:

$$\overline{S_{\text{vN}}[\rho_R]} = \lim_{n\to 1} \frac{1}{1-n}\left(\overline{\ln Z(B_n)} - n\overline{\ln Z(B)}\right)$$
$$= \lim_{n\to 1} \frac{1}{1-n}\left(\lim_{m\to 0} \frac{1}{m}\left(\mathcal{P}(B_n^m)-1\right) - n \lim_{m\to 0}\frac{1}{m}\left(\mathcal{P}(B^m)-1\right)\right)$$



The $\overline{\ln Z}$ Replica Trick



We just need this:

The $\overline{\ln Z}$ Replica Trick

And since we're working with AdS boundary conditions (e.g. unlike the East Coast calculation we don't care about the flat region)

Replica Wormholes in $\overline{\ln Z}$

Is there a simple theory of gravity with a regime where replica wormholes actually contribute to $\overline{\ln Z}$ at least as much as the disconnected topologies?

In JT gravity (and also some other cases), there have been numerous studies of the lack of factorization due to contributions of connected topologies $\mathcal{P}(B^m)$. Okuyama, Sakai; Johnson; Okuyama...

So obviously these connected topologies should contribute to

$$\frac{1}{m}\left(\mathcal{P}(B^m) - 1\right)$$

But the $m \to 0$ analytic continuation is subtle. And as it turns out, something much more interesting happens.

Table of Contents

Big Picture: what do we see NE, Fischetti, Maloney

- We do find that replica wormholes give a larger contribution to $\overline{\ln Z}$ than disconnected topologies at low temperatures.
- In fact, the disconnected topologies result in a pathological free energy. So it is clear that this is not the full story.
- The pathological behavior prima facie appears to be mitigated by the inclusion of replica wormholes, but in fact this is not sufficient.
- Even with the inclusion of replica wormholes (and a resummation of genus), the free energy is non-monotonic with temperature.
- The only thing that can possibly go wrong is the analytic continuation itself, $m \to 0$. There is a lot of freedom in how to do this, and the obvious/simple analytic continuations are clearly wrong.
- This is highly reminiscent of replica symmetry breaking in spin glass systems, which require a similar computation of the quenched free energy, where the naive analytic continuation of the replica wormholes ameliorates but does not fix the problems in the free energy. This parallel suggest that the correct analytic continuation requires a version of replica symmetry breaking.

Big Picture: what do we see

I sadly don't have time to go through the calculations in technical detail today, so I will distill the basic results into a few slides. Technical questions are welcomed on slack or email!

I also do not have the time to go through the fascinating parallels with spin glasses, so I will most skip over them really I only mentioned them to get you to read the paper.

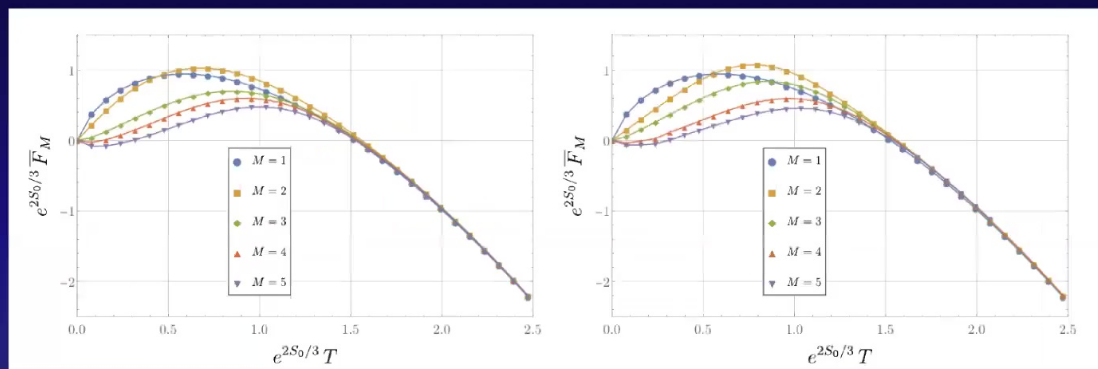
The big takeaway is that in at least some (low-dimensional) theories of gravity, there exists at least one regime in which replica wormholes must make a large contribution to $\overline{\ln Z}$ to avoid various pathologies, and this contribution requires a highly nontrivial analytic continuation that appears highly analogous to replica symmetry breaking in spin glasses.

The Quenched Free Energy

... so that the quenched free energy is

$$\overline{F} = -T \lim_{M \to \infty} \lim_{m \to 0} \frac{1}{m} \left(\mathcal{P}_{m,M} - 1 \right)$$

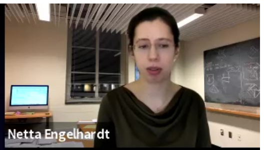
For a given M, we can analytically continue to $m \to 0$.



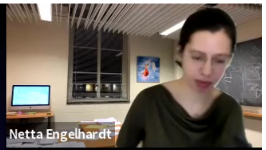
In fact, the analytic continuation to $m \to 0$ is not unique. Here are two different analytic continuations: qualitatively similar but quantitatively different.

We know that the $\overline{F}$ is not correct because we chose the wrong analytic continuation. But we also have learned that replica wormholes not only *must* contribute, but they must do so in a highly nontrivial way via some analytic continuation that, without some top-down derivation, we will have to try and guess.

Overview



- The recent developments on the BH info front have suggested that Euclidean wormholes really should be included in the gravitational path integral. One way of interpreting this (though not the only one) is as computing an ensemble average.
- Ultimately we'd like to understand how general this statement is (higher dimensions?)
- We'd like to know how to calculate everything the GPI computes without using the GPI black box
- All of which means understanding the GPI better



Summary

- Here we are interested in probing the replica wormhole phenomenon in the GPI by looking at an arguably more primitive quantity than the von Neumann entropy: $\ln Z$.
- In JT we find that at sufficiently low temperatures (but not so low that the genus expansion is out of control), the free energy without replica wormholes is severely pathological
- It's also pathological with replica wormholes.
- This is very similar to spin glasses, where a novel analytic continuation removes the pathologies altogether.