Replica Wormholes and Holographic Entanglement Negativity

Xi Dong

UC SANTA BARBARA

October 25, 2021

BIRS Workshop on "Gravitational Emergence in AdS/CFT"

This talk is

- based on work with Sean McBride and Wayne Weng (just appeared) [2110.11947]

- related to previous work with Xiaoliang Qi and Michael Walter [2101.11029]

We will learn:

- What is negativity?
- Why negativity?
- Negativity in JT gravity with end-of-the-world (EOW) branes.
- Four phases for negativity.
- Replica wormholes that break the replica symmetry spontaneously.
- Resolvent, Schwinger-Dyson equation, and negativity spectrum near phase transitions.

What is negativity?

Negativity is a measure of quantum entanglement in mixed states.
Definition:

- Given $\rho_{R_1 R_2}$ on $\mathcal{H}_{R_1} \otimes \mathcal{H}_{R_2}$, choose an orthonormal basis $|i\rangle$ for R_1 and an orthonormal basis $|j\rangle$ for R_2.
- Partial transpose $\rho_{R_1 R_2}^{T_2}$ on R_2:

$$\langle ij|\rho_{R_1 R_2}^{T_2}|i'j'\rangle \equiv \langle ij'|\rho_{R_1 R_2}|i'j\rangle$$

- Its eigenvalues λ_i are real (and sum to one), but can be negative.
- How negative?

$$\text{Negativity: } \mathcal{N} \equiv \sum_i \frac{|\lambda_i| - \lambda_i}{2}$$

$$\text{Logarithmic negativity: } \mathcal{E} \equiv \log\left(\sum_i |\lambda_i|\right) = \log(2\mathcal{N} + 1)$$

$$\mathcal{N} \equiv \sum_i \frac{|\lambda_i| - \lambda_i}{2}, \quad \mathcal{E} \equiv \log\left(\sum_i |\lambda_i|\right)$$

- Both vanish if all $\lambda_i \geq 0$.
- This is true if e.g. $\rho_{R_1R_2}$ is a separable (unentangled) state:

$$\rho_{R_1R_2} = \sum_k p_k \rho_{R_1}^{(k)} \otimes \rho_{R_2}^{(k)}, \qquad \sum_k p_k = 1$$

- How about an entangled state? Consider an EPR pair:

$$\rho_{R_1R_2} = \frac{1}{2}(|00\rangle + |11\rangle)(\langle 00| + \langle 11|) = \frac{1}{2}\begin{array}{c} \begin{matrix} 00 & 01 & 10 & 11 \end{matrix} \\ \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix} \end{array} \begin{matrix} 00 \\ 01 \\ 10 \\ 11 \end{matrix}$$

$$\Rightarrow \quad \rho_{R_1R_2}^{T_2} = \frac{1}{2}\begin{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}^T & \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}^T \\ \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}^T & \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}^T \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$\Rightarrow \quad \lambda_i = \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, -\frac{1}{2} \quad \Rightarrow \quad \mathcal{N} = \frac{1}{2}, \quad \mathcal{E} = \log 2$$

- In general, logarithmic negativity $\geq$ distillable entanglement.

- Negativity is an entanglement monotone – does not increase under any LOCC (Local Operations and Classical Communication).

Renyi negativity

Similar to the generalization of von Neumann entropy to Renyi entropy:

$$\text{Renyi negativity:}\ \ N_n \equiv \mathrm{Tr}(\rho_{R_1 R_2}^{T_2})^n$$

Analytic continuations from integer to real n are different for even and odd cases (as λ_i can be negative):

$$N_{2m}^{(\text{even})} = \sum_i |\lambda_i|^{2m}$$
$$N_{2m-1}^{(\text{odd})} = \sum_i \mathrm{sgn}(\lambda_i)|\lambda_i|^{2m-1}$$

Recovers logarithmic negativity $\mathcal{E}$ (and $\mathcal{N}$):

$$\mathcal{E} \equiv \log\left(\sum_i |\lambda_i|\right) = \lim_{m\to 1/2} \log N_{2m}^{(\text{even})}$$

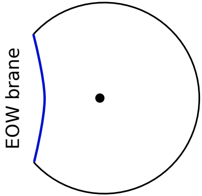
Why negativity?

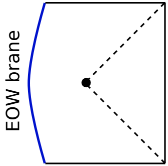
- Negativity and its Renyi versions provide useful (and relatively tractable) measures of multipartite entanglement (R_1, R_2, $\overline{R_1 R_2}$).
- In general quantum states, the structure of multipartite entanglement is very rich and not well-understood.
- Holographic states have highly constrained – and often more tractable – structure of multipartite entanglement.
- It is thus important to understand holographic negativity to gain insight on gravitational questions such as the black hole information problem.
- This is particularly true in light of recent work on understanding the von Neumann entropy of Hawking radiation from replica wormholes, as we need more detailed information about the structure of its quantum state to fully understand the dynamics of black hole evaporation.

A generalization of the 2d gravity model in [Penington, Shenker, Stanford, Yang].

- Consider a black hole (BH) in Jackiw-Teitelboim (JT) gravity with an EOW brane behind the horizon:

$$I_{JT} = -\frac{S_0}{4\pi}\int \sqrt{g}R - \frac{1}{2}\int \sqrt{g}\phi(R+2) + I_{\text{bdy}}$$





- The EOW brane has a large number k of internal states.
- To model an evaporating black hole B, entangle it with an auxiliary system R modeling the early radiation:

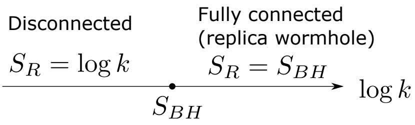
$$|\Psi\rangle = \frac{1}{\sqrt{k}}\sum_{i=1}^{k}|\psi_i\rangle_B|i\rangle_R$$

Entropy of the radiation

The entropy of the radiation follows the Page curve: [Penington, Shenker, Stanford, Yang]

$$S_R = \min\{\log(k), S_{BH}\}$$

Phases:



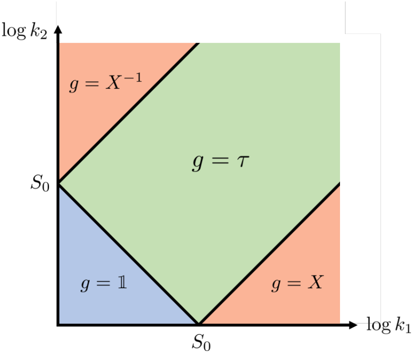
Negativity of the radiation

Now divide the radiation R into two subsystems R_1, R_2:

$$\mathcal{H}_R = \mathcal{H}_{R_1} \otimes \mathcal{H}_{R_2}, \qquad k = k_1 \times k_2.$$

To measure the entanglement between R_1 and R_2 in the mixed state $\rho_{R_1 R_2}$, we now study the negativities: $\mathcal{N}$, $\mathcal{E}$, N_n.

✓ Leads to a rich phase diagram:



To see this, use the 2d gravitational path integral to calculate the Renyi negativity of $\rho_{R_1 R_2}$.

- The natural generalization of $|\Psi\rangle = \frac{1}{\sqrt{k}} \sum_{i=1}^{k} |\psi_i\rangle_B |i\rangle_R$ is

$$|\Psi\rangle = \frac{1}{\sqrt{k}} \sum_{i=1}^{k_1} \sum_{j=1}^{k_2} |\psi_{ij}\rangle_B |ij\rangle_{R_1 R_2}$$

  $|ij\rangle_{R_1 R_2} \equiv |i\rangle_{R_1} |j\rangle_{R_2}$ is an orthonormal basis of R.
- The EOW brane carries two flavor indices i, j.
- The density matrix

$$\rho_{R_1 R_2} = \frac{1}{k} \sum_{i,i'=1}^{k_1} \sum_{j,j'=1}^{k_2} |i'j'\rangle\langle ij|_{R_1 R_2} \langle \psi_{ij} | \psi_{i'j'} \rangle_B$$

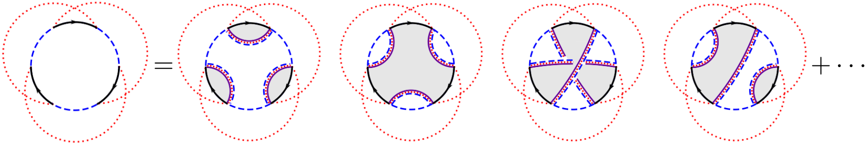
- Partial transpose on R_2

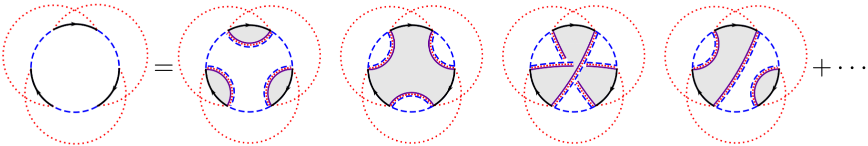
$$\rho_{R_1 R_2}^{T_2} = \frac{1}{k} \sum_{i,i'=1}^{k_1} \sum_{j,j'=1}^{k_2} |i'j'\rangle\langle ij|_{R_1 R_2} \langle \psi_{ij'} | \psi_{i'j} \rangle_B$$

$$\rho_{R_1 R_2}^{T_2} = \frac{1}{k} \sum_{i,i'=1}^{k_1} \sum_{j,j'=1}^{k_2} |i'j'\rangle\langle ij|_{R_1 R_2} \langle \psi_{ij'} | \psi_{i'j} \rangle_B$$

The matrix element $\langle i'j'|\rho_{R_1 R_2}^{T_2}|ij\rangle \propto \langle \psi_{ij'}|\psi_{i'j}\rangle_B$ is found by a gravitational path integral with boundary condition

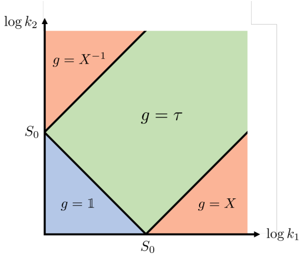
Renyi negativity $N_n \equiv \mathrm{Tr}(\rho_{R_1 R_2}^{T_2})^n$ is found by (e.g., $n = 3$)



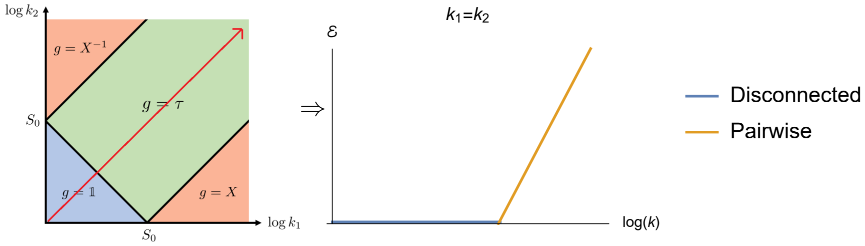


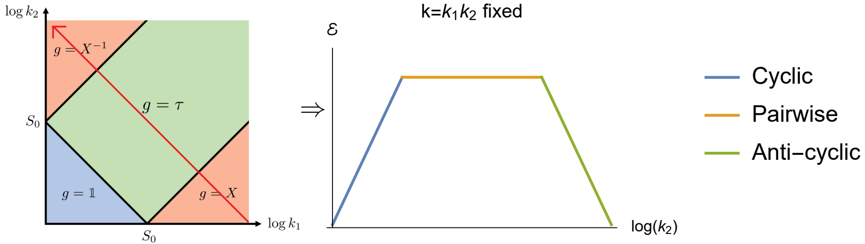
1. Disconnected geometry: $k_1 k_2 Z_1^3 \sim k_1 k_2 e^{3S_0}$
2. Cyclically connected wormhole: $k_1^3 k_2 Z_3 \sim k_1^3 k_2 e^{S_0}$
3. 'Anti-cyclically' connected wormhole: $k_1 k_2^3 Z_3 \sim k_1 k_2^3 e^{S_0}$
4. Pairwise connected wormhole: $k_1^2 k_2^2 Z_1 Z_2 \sim k_1^2 k_2^2 e^{2S_0}$

- If $k \ll e^{S_0}$, disconnected geometry dominates.
- If $k \gg e^{S_0}$, the other three compete:
  - If $k_1 \gg k_2 e^{S_0}$, cyclic geometry dominates.
  - If $k_2 \gg k_1 e^{S_0}$, anti-cyclic geometry dominates.
  - If $e^{-S_0} \ll \frac{k_1}{k_2} \ll e^{S_0}$, pairwise geometries dominate, spontaneously breaking the replica symmetry. (A compromise between cyclic and anti-cyclic geometries.)



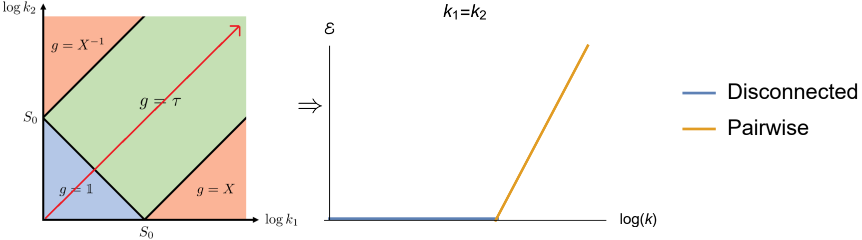
Negativity in four phases



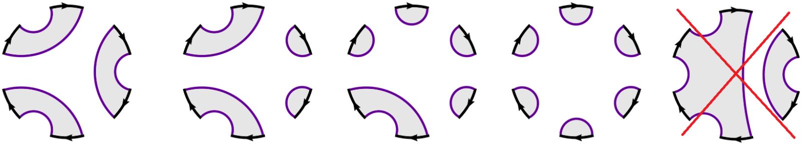


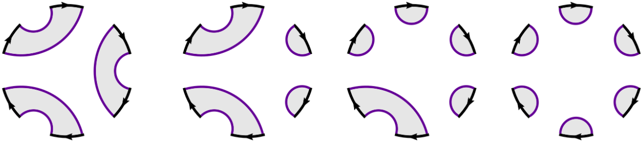
Agrees with general holographic settings studied in [XD, Qi, Walter].

Near the disconnected-pairwise transitions



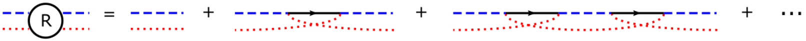
- Here, both disconnected and pairwise geometries could dominate, along with all other geometries made of single-boundary disks and pairwise connected wormholes:



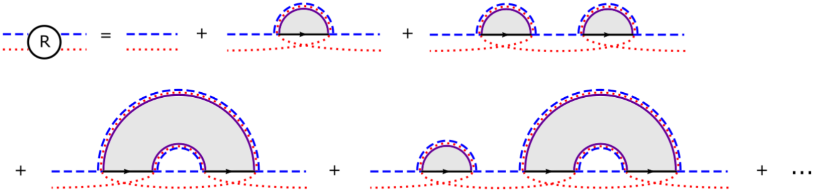


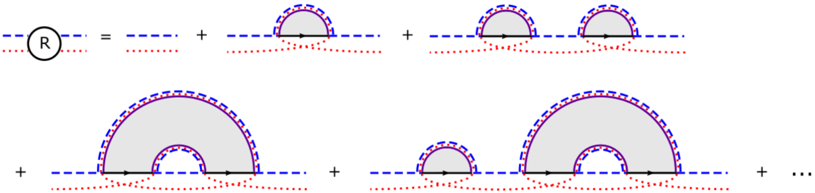
- Renyi negativity is found by summing these geometries.
- To do this, define the resolvent matrix of $\rho_{R_1 R_2}^{T_2}$:

$$R_{ij,i'j'}(\lambda) = \left(\frac{1}{\lambda \mathbb{1} - \rho_{R_1 R_2}^{T_2}} \right)_{ij,i'j'} = \sum_{n=0}^{\infty} \frac{1}{\lambda^{n+1}} \left[(\rho_{R_1 R_2}^{T_2})^n \right]_{ij,i'j'}$$

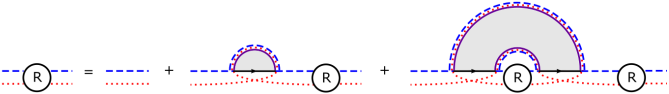


- In the bulk, sum over geometries with any number of boundaries:





- Can be summed by a Schwinger-Dyson equation:



It indeed generates the infinite sum over geometries made of single-boundary and pairwise components (on the top of this slide).

- Taking the trace $R = \sum_{i=1}^{k_1} \sum_{j=1}^{k_2} R_{ij,ij}$:

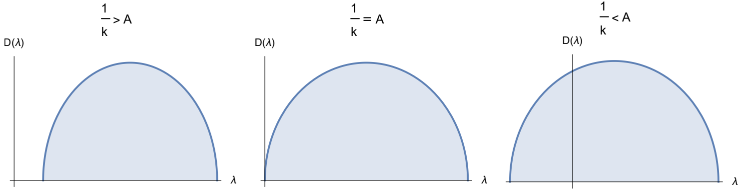
$$\lambda R = k + \frac{R}{k} + \frac{Z_2 R^2}{(kZ_1)^2}$$

$$\lambda R = k + \frac{R}{k} + \frac{Z_2 R^2}{(kZ_1)^2}$$

Negativity spectrum: eigenvalue density $D(\lambda) = \frac{1}{2\pi i}\left[R(\lambda - i\epsilon) - R(\lambda + i\epsilon)\right]$

$$D(\lambda) = \frac{2k}{\pi A^2}\sqrt{A^2 - \left(\lambda - \frac{1}{k}\right)^2}, \qquad A \equiv \sqrt{\frac{4Z_2}{kZ_1^2}} \sim \frac{1}{\sqrt{ke^{S_0}}}$$

Is the Wigner semicircle distribution on $[\frac{1}{k} - A, \frac{1}{k} + A]$:

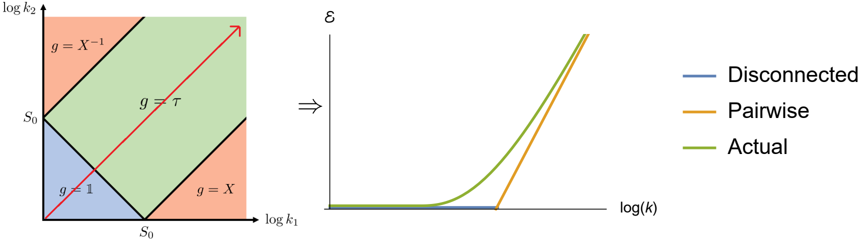


- Negativity turns on at $\frac{1}{k} = A$ (or $k \sim e^{S_0}$)
- All Renyi negativities can be calculated analytically.

For example, logarithmic negativity:

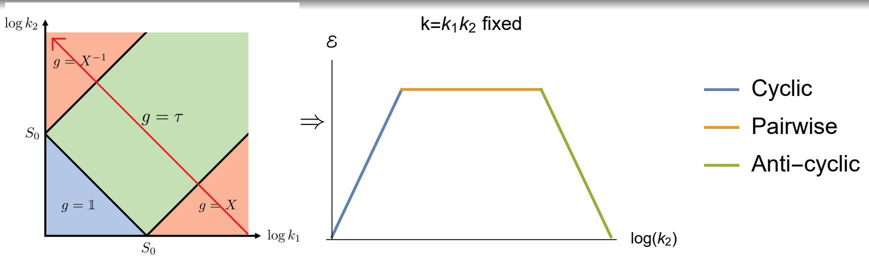
$$\mathcal{E} \equiv \log \sum_i |\lambda_i| = \log \int d\lambda\, D(\lambda)|\lambda|$$
$$= \log \left[\frac{2}{\pi} \left(\frac{\sqrt{k^2A^2 - 1}(2k^2A^2 + 1)}{3k^2A^2} + \operatorname{arccsc}(kA) \right) \right] I_{kA>1}$$

Smooth out the sharp transition:

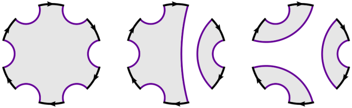


⇒ $\mathcal{O}(1)$ correction at this disconnected-pairwise transition.

Near the cyclic-pairwise transition

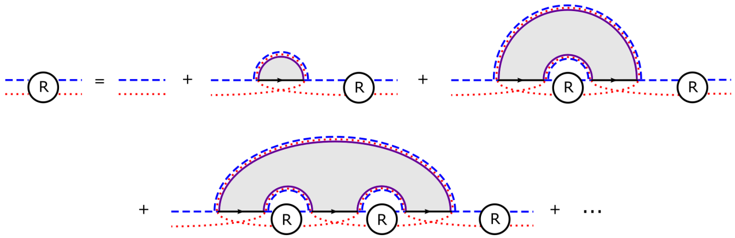


- Here, the dominant geometries include the cyclic and pairwise wormholes, as well as more generic wormholes that connect more than 2 boundaries:



- Let us simply sum over all planar geometries.
- Doing so is in fact valid in a larger regime than we need.
- Valid anywhere away from the anti-cyclic phase, including the disconnected-pairwise transition and the (cyclic) triple point.
- The price we pay: it is more difficult.
- But if done, the anti-cyclic-pairwise transition follows by $k_1 \leftrightarrow k_2$.

Sum planar geometries by a Schwinger-Dyson equation (now with higher-order terms):



$$\lambda R = k + \frac{R}{k} + \frac{Z_2 R^2}{(kZ_1)^2} + k_2 \sum_{n\geq 3,\ \text{odd}} \frac{Z_n}{k_2^n} \frac{R^n}{(kZ_1)^n} + k_2^2 \sum_{n\geq 4,\ \text{even}} \frac{Z_n}{k_2^n} \frac{R^n}{(kZ_1)^n}$$

- Setting $k_2 = 1$ recovers the result of [Penington, Shenker, Stanford, Yang].
- Can be solved exactly if black hole is in a microcanonical ensemble.
- Can be solved approximately for a canonical ensemble, leading to enhanced corrections at the phase transition.

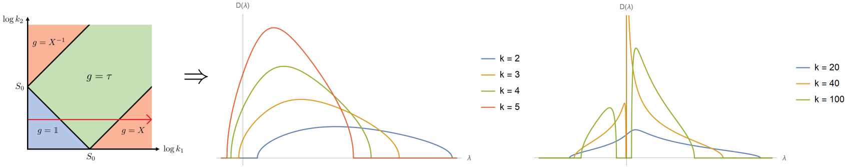
Microcanonical ensemble

$$\lambda R = k + \frac{R}{k} + \frac{Z_2 R^2}{(kZ_1)^2} + k_2 \sum_{n\geq 3,\ \text{odd}} \frac{Z_n}{k_2^n} \frac{R^n}{(kZ_1)^n} + k_2^2 \sum_{n\geq 4,\ \text{even}} \frac{Z_n}{k_2^n} \frac{R^n}{(kZ_1)^n}$$

- In a microcanonical ensemble, $Z_n = e^{\mathbf{S}} (\#)^n$.
- The resolvent equation can be resummed:

$$\lambda R = k + \frac{R}{k} \frac{1 + e^{-\mathbf{S}} R/k}{1 - (e^{-\mathbf{S}} R/kk_2)^2}$$

- Solving this cubic equation gives the exact negativity spectrum:



Start: Wigner semicircle distribution (in disconnected and pairwise phase)
$\Rightarrow$ develops singularities (at pairwise-cyclic transition).
$\Rightarrow$ two disjoint Marchenko-Pastur distributions (in cyclic phase).

A topological aside

- Instead of JT gravity, we could also study a topological model of 2d gravity [Marolf, Maxfield]:

$$S_{\text{top}} = -S_0 \chi(M) - S_\partial |\partial M|$$

- Repeating our analysis gives a cubic equation very similar to

$$\lambda R = k + \frac{R}{k} \frac{1 + e^{-\mathbf{S}} R/k}{1 - (e^{-\mathbf{S}} R/kk_2)^2}$$

- The topological model is analogous to the microcanonical JT model.
- Both models are in fact similar to negativity in random mixed states [Shapourian, Liu, Kudler-Flam, Vishwanath; Kudler-Flam, Narovlansky, Ryu].

Back to JT: canonical ensemble

$$\lambda R = k + \frac{R}{k} + \frac{Z_2 R^2}{(kZ_1)^2} + k_2 \sum_{n\geq 3,\ \text{odd}} \frac{Z_n}{k_2^n} \frac{R^n}{(kZ_1)^n} + k_2^2 \sum_{n\geq 4,\ \text{even}} \frac{Z_n}{k_2^n} \frac{R^n}{(kZ_1)^n}$$

- In a canonical ensemble, $Z_n = e^{S_0} \int_0^\infty ds \rho(s) y(s)^n$, where

$$\rho(s) = \frac{s}{2\pi^2} \sinh(2\pi s), \quad y(s) \equiv e^{-\frac{\beta s^2}{2}} 2^{1-2\mu} \left| \Gamma\left(\mu - \frac{1}{2} + is\right) \right|^2$$

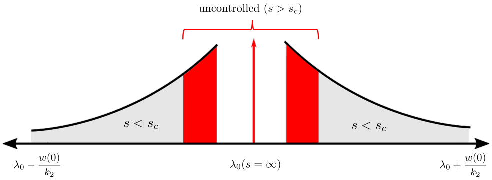
- The resolvent equation can be resummed inside the s-integral:

$$\lambda R = k + k_2^2 e^{S_0} \int_0^\infty ds \rho(s) \frac{w(s) R (k + w(s) R)}{k^2 k_2^2 - w(s)^2 R^2}.$$

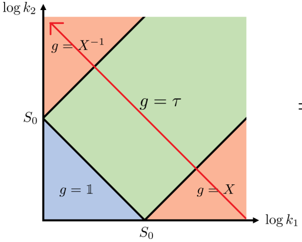
- Can be solved (in a complicated way) in the semiclassical limit ($\beta \to 0$).

Results near the cyclic-pairwise transition

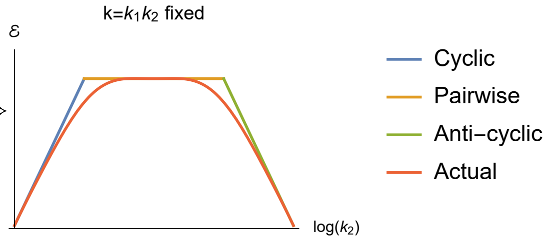
Negativity spectrum:



- Eigenvalues in the uncontrolled region cause a very small error.
- Smooth out the sharp transition in e.g. logarithmic negativity:



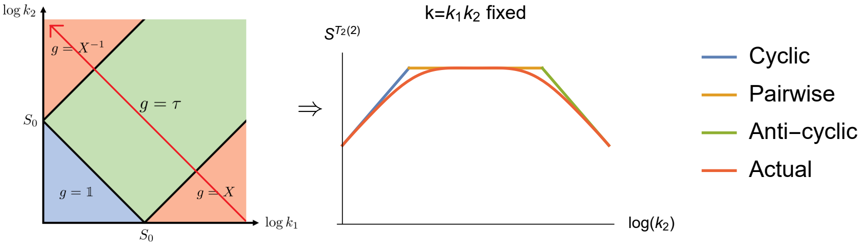
$\Rightarrow$



$\Rightarrow \Delta\mathcal{E} \equiv \mathcal{E}_{\text{actual}} - \mathcal{E}_{\text{naive}} = -\frac{\pi^2}{8\beta}$: enhanced correction from infinite sum of saddles (similar to $\mathcal{O}(1/\beta)$ correction to S_n with $n < 1$ at Page transition)

Results near the cyclic-pairwise transition

Mostly similar behaviors in other Renyi negativities, but one stands out:
Refined Rényi-2 negativity: $S^{T_2(2)} \equiv -\sum_i \overline{\lambda_i^2} \log \overline{\lambda_i^2}, \quad \overline{\lambda_i^2} \equiv \frac{\lambda_i^2}{\sum_j \lambda_j^2}.$



$$\Rightarrow \Delta S^{T_2(2)} \equiv S^{T_2(2)}_{\text{actual}} - S^{T_2(2)}_{\text{naive}} = -2\sqrt{\frac{\pi}{\beta}}$$

- This enhanced correction is similar to the $\mathcal{O}(1/\beta)$ correction to the von Neumann entropy at the Page transition [Penington, Shenker, Stanford, Yang; XD, Wang; Marolf, Wang, Wang].
- This is perhaps not surprising once we recall holographically $S^{T_2(2)}$ can be written as the sum of von Neumann entropies [XD, Qi, Walter].

Questions

- What is the interpretation of the $\mathcal{O}(1/\beta)$ enhanced correction to the logarithmic negativity at the phase transition? Can it be explained in a way similar to the diagonal approximation (for von Neumann)?
- Can we connect these results to other entanglement measures, such as the reflected entropy and entanglement of purification? What does the phase diagram look like for them?
- Can we generalize our analysis to other toy models of evaporating black holes, such as the examples in [Almheiri, Hartman, Maldacena, Shaghoulian, Tajdini].
- How can we use these results to understand holography as a very special type of quantum error-correcting code, or as some improved type of tensor network?

Thank you.