ETH zürich

Demystifying the replica trick calculation of the entropy of Hawking radiation

Jinzhao Wang

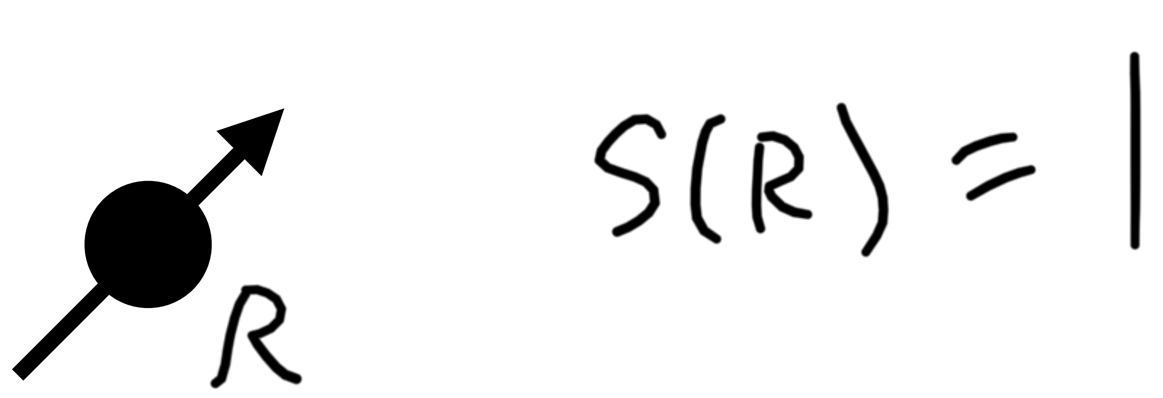
Loops'22

18.07.2022

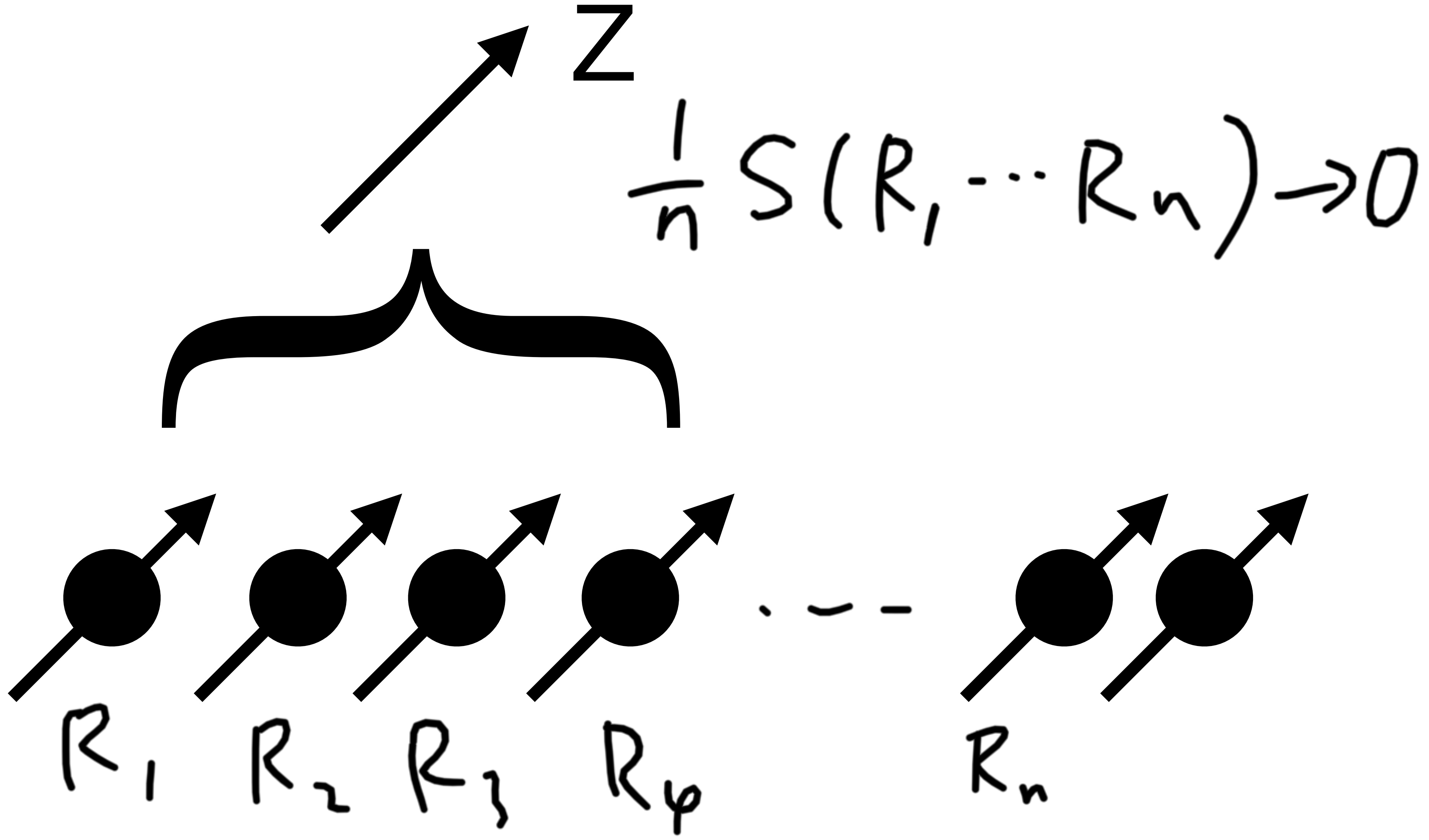
Joint work with Renato Renner, arXiv: 2110.14653

A motivating example

- How do we describe the state of a spin-1/2 particle (prepared in the +Z direction) without a reference frame? One (or perhaps the best) option is to assign a maximally mixed state, with maximal entropy.

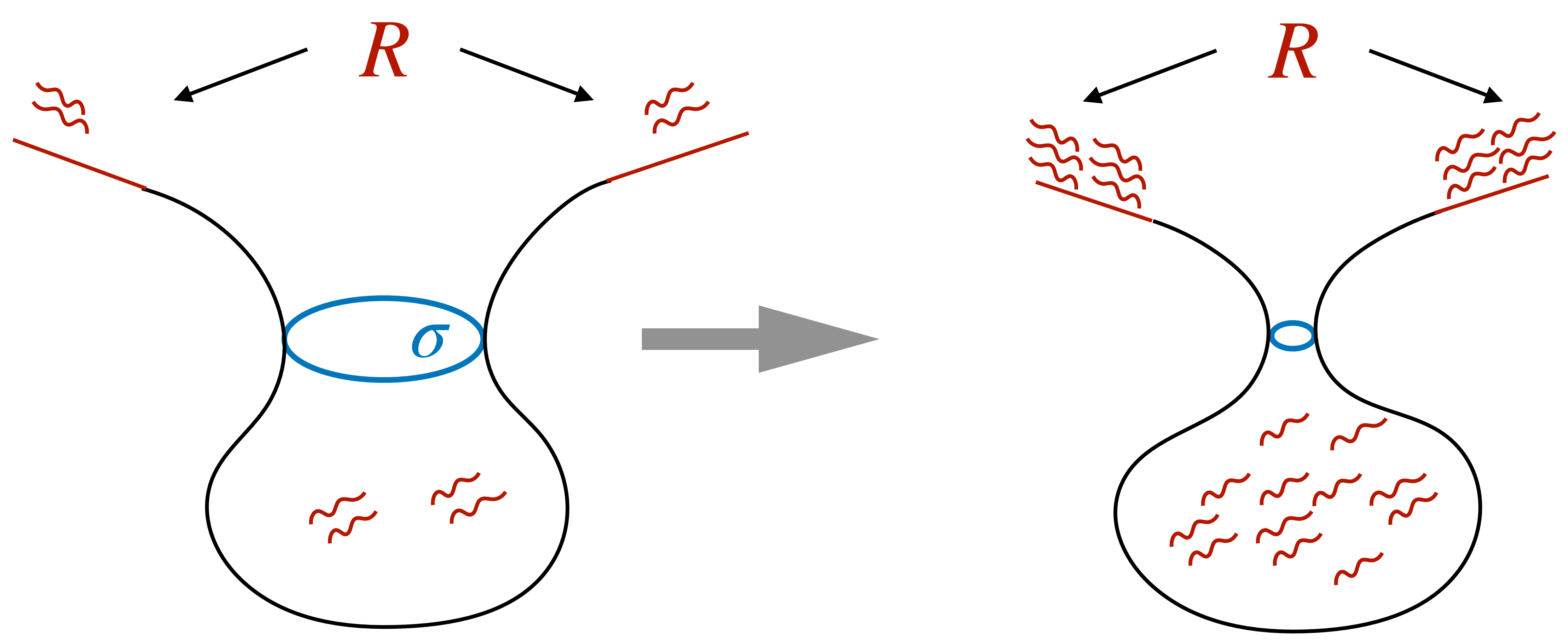


- However, with many copies we can apply measures to a sample of copies to figure out the spin direction, so we can meaningfully assign $|0\rangle$ to each of the remaining particles, which have zero entropy. (The sample serves as a reference for the rest.)



This is the essence of the black hole information puzzle, the semiclassical gravity lacks a "reference" in its description of a black hole spacetime.

The black hole information puzzle

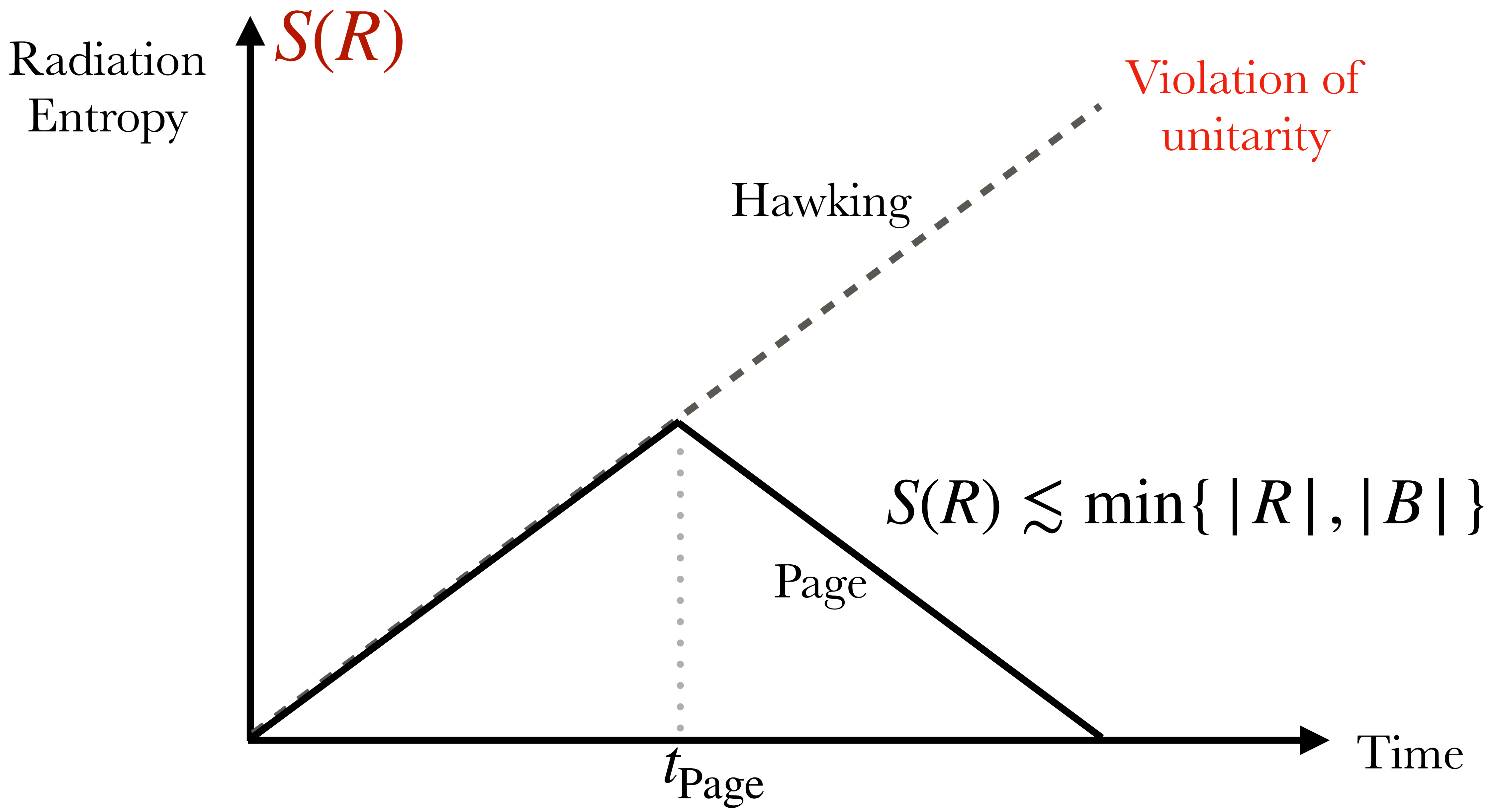


$$\rho_R \approx \rho_{\text{thermal}} \approx \frac{I}{2^k}$$

If the radiation is thermal and featureless as shown by Hawking, the radiation entropy, $S(R) = -\operatorname{Tr}\rho_R \log \rho_R$, shall keep increasing. Where is the information carried by the collapsing star after the black hole evaporation?

[Hawking]

How would the entropy behave if we demand unitarity?



If black hole were to evolve unitarily as demanded by quantum theory, say under some typical random unitary, then the radiation entropy should follow the **Page curve**.

[Page]

The challenge is to obtain the Page curve using gravity calculations and understand its discrepancy from Hawking's calculation.

Recent progress: replica trick and replica wormholes

$S(R) = -\,\mathrm{Tr}\rho_R \log \rho_R$ gives us the non-unitary answer, so let's try a different way to compute the radiation entropy: [Cardy, Calabrese]

$$S(R) = \lim_{n\to 1} \frac{1}{1-n} \log \mathrm{Tr}\rho_R^n = \lim_{n\to 1} \frac{1}{1-n} \log \mathrm{Tr}\rho_R^{\otimes n}\tau_n$$

replicas

cyclic shift operator

One can use the gravitational path integral (GPI) to capture $\mathrm{Tr}\rho_R^{\otimes n}\tau_n$ by setting up appropriate boundary conditions.

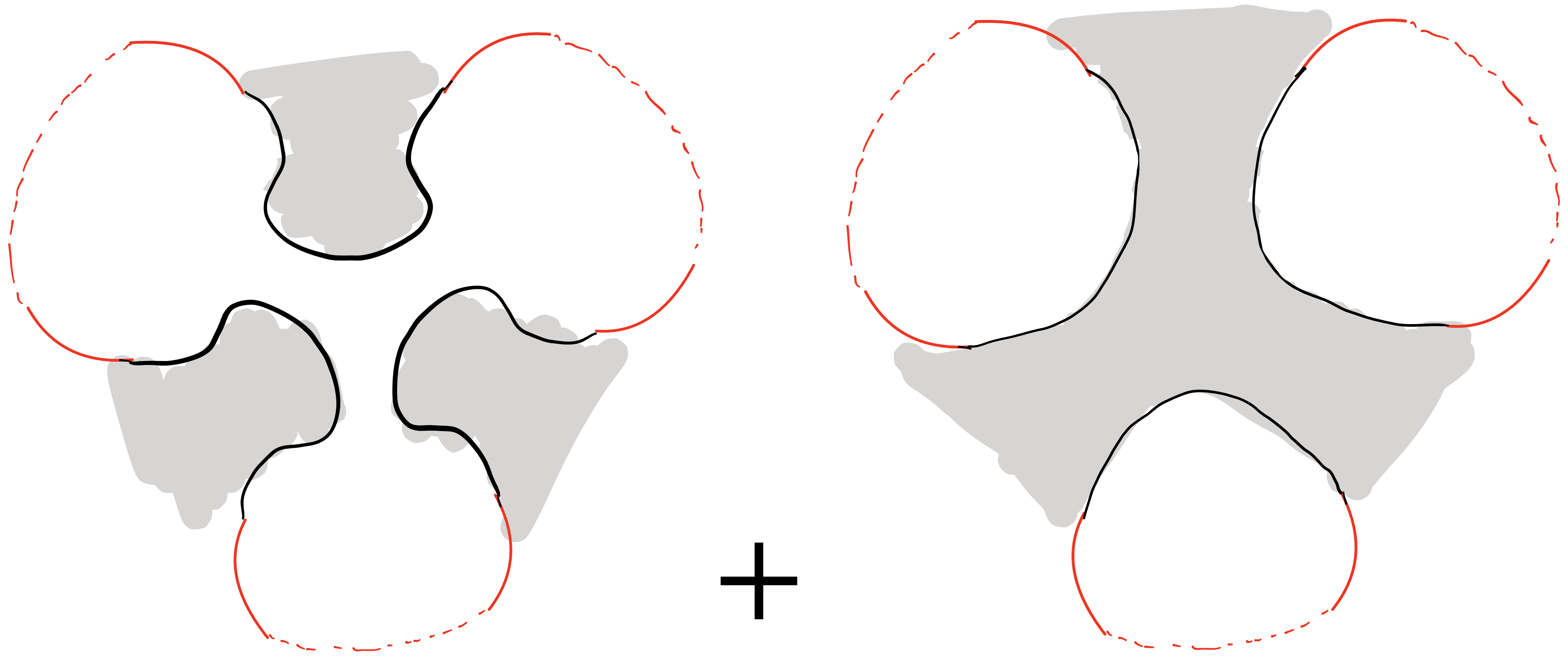
$$S^{\mathrm{grav}}(R) := \lim_{n\to 1} \frac{1}{1-n} \log\langle \tau_n \rangle := \lim_{n\to 1} \frac{1}{1-n} \log \frac{Z[\mathscr{B}^{\times n}, \tau_n]}{Z[\mathscr{B}]^n} \,.$$

The full GPI is tricky to evaluate, so we use the saddlepoint approximation. A fully disconnected Hawking saddle for the replica trick GPI gives Hawking's answer, but there is another important saddle known as the **replica wormholes** .

[Hawking, Gibbons, Hartle, Lewkowycz, Maldacena, Faulkner]

$Z[\mathscr{B}^{\times 3}, \tau_3]$

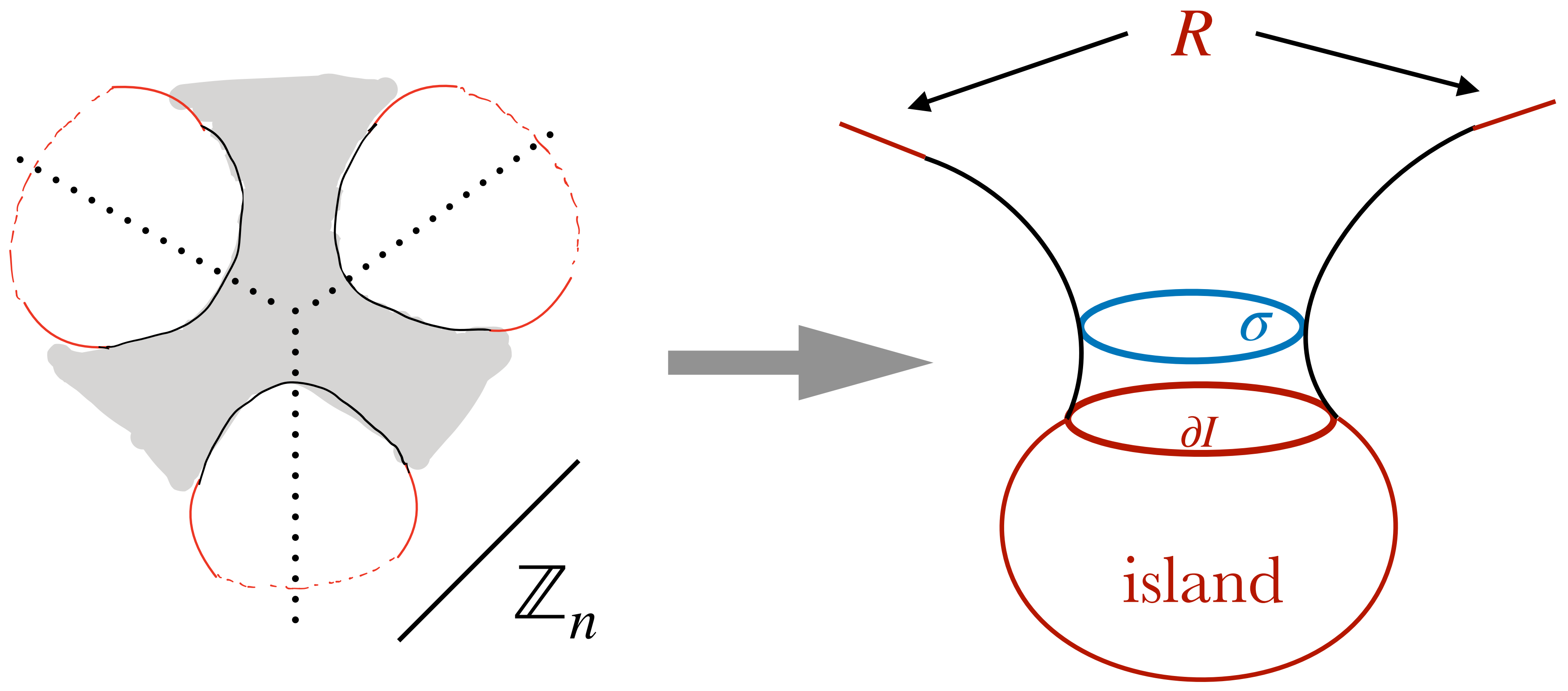
boundary conditions: dashed lines indicate τ_3, solid lines (without the dashed lines contracting them) represent $\rho_R^{\otimes n}$



the Hawking saddle with disconnected black holes

the wormhole saddle with a connected replica wormhole

The Page curve from the island formula



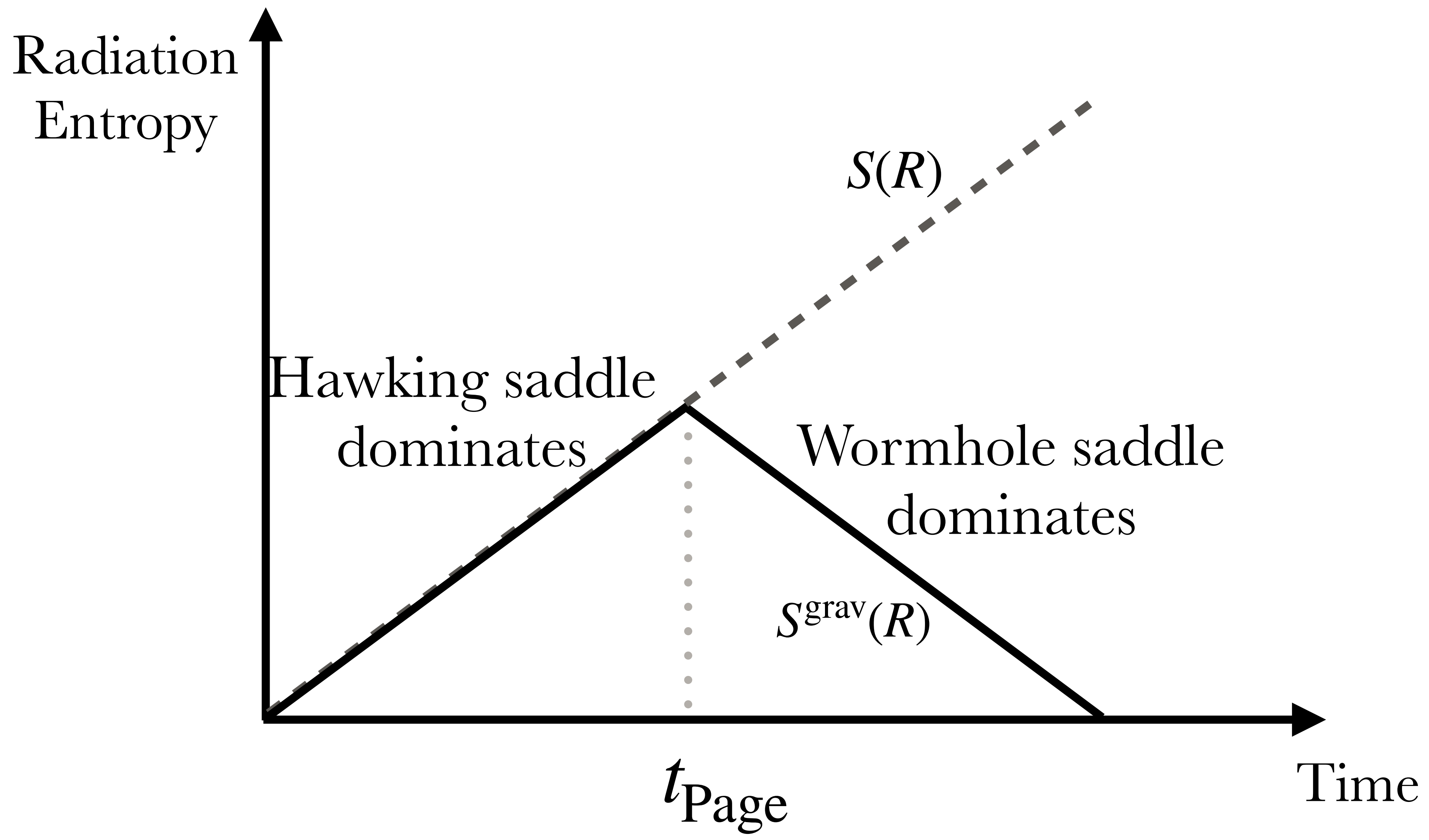
Islands are gateways to replica wormholes

"Folding up" replica wormholes in the $n \to 1$ limit leads to the **island formula**:

$$S^{\text{grav}}(R) := \min \text{ext}_I S_{\text{gen}}[I] = \min \text{ext}_I \left(\frac{A[\partial I]}{4G_N} + S(I \cup R) \right).$$

By changing the surface to evaluate the generalized entropy, we obtain the Page curve.

[Engelhardt, Wall, Almheiri, Maldacena, Mahajan, Zhao, Hartman, Shaghoulian, Tajdini, Penington, Shenker, Stanford, Yang,...]



For a young black hole (w.r.t. t_{Page}), the island is empty (with the Hawking saddle dominating); and the island forms (with the replica wormhole dominating) in the interior for an old black hole, which bends the curve down.

Some issues with the island formula

$$S^{\mathrm{grav}}(R) := \min \mathrm{ext}_I S_{\mathrm{gen}}[I] = \min \mathrm{ext}_I \left(\frac{A[\partial I]}{4G_N} + S(I \cup R) \right).$$

- The radiation system "R" appears on both sides but interpreted differently.
- Hawking's prediction is "incorrect" in an uncannily precise way such that it is still an essential input to the island formula oracle.
- What exactly is $S^{\mathrm{grav}}(R)$ in terms of familiar entropy measures in quantum information? Why is it different from Hawking's $S(R)$?

Using the quantum de Finetti theorem, we will clarify $S^{\mathrm{grav}}(R)$
to address these questions.

Demystifying the replica trick with the Quantum de Finetti theorem

$$S(R) := \lim_{n\to 1} \frac{1}{1-n} \log \mathrm{Tr}\rho_R^n = \lim_{n\to 1} \frac{1}{1-n} \log \mathrm{Tr}\rho_R^{\otimes n}\tau_n$$

$$S^{\mathrm{grav}}(R) := \lim_{n\to 1} \frac{1}{1-n} \log\langle\tau_n\rangle = \lim_{n\to 1} \frac{1}{1-n} \log \frac{Z[\mathscr{B}^{\times n}, \tau_n]}{Z[\mathscr{B}]^n}$$

equality holds if $\langle\tau_n\rangle = \mathrm{Tr}\rho_R^{\otimes n}\tau_n$

Apparently, Hawking "forgot" the replica wormholes, but the discrepancy $S^{\mathrm{grav}}(R) \neq S(R)$ only indicates that the replicas are **correlated**, such that joint state $\rho_{R_1\cdots R_n}$ of the radiation is correlated,

$$\langle\tau_n\rangle = \mathrm{Tr}\rho_{R_1\cdots R_n}\tau_n \neq \mathrm{Tr}\rho_R^{\otimes n}\tau_n\,.$$

Since the state is prepared with identical and independent boundary conditions, the state is nonetheless **invariant under any permutations of the replicas**. The **quantum de Finetti theorem** then says: the permutation-invariance of N replicas implies the any subset of n-partite state is a convex combination of product states over some **ensemble** $\mathscr{W}$,

$$\rho_{R_1\cdots R_n} = \sum_{w\in\mathscr{W}} p_w\, \rho_{R|w}^{\otimes n}\,.$$

[de Finetti, Caves, Fuchs, Schack, Koenig, Renner, Christandl, Mitchison,...]

Viewpoint 1: The conditional entropy of Hawking radiation

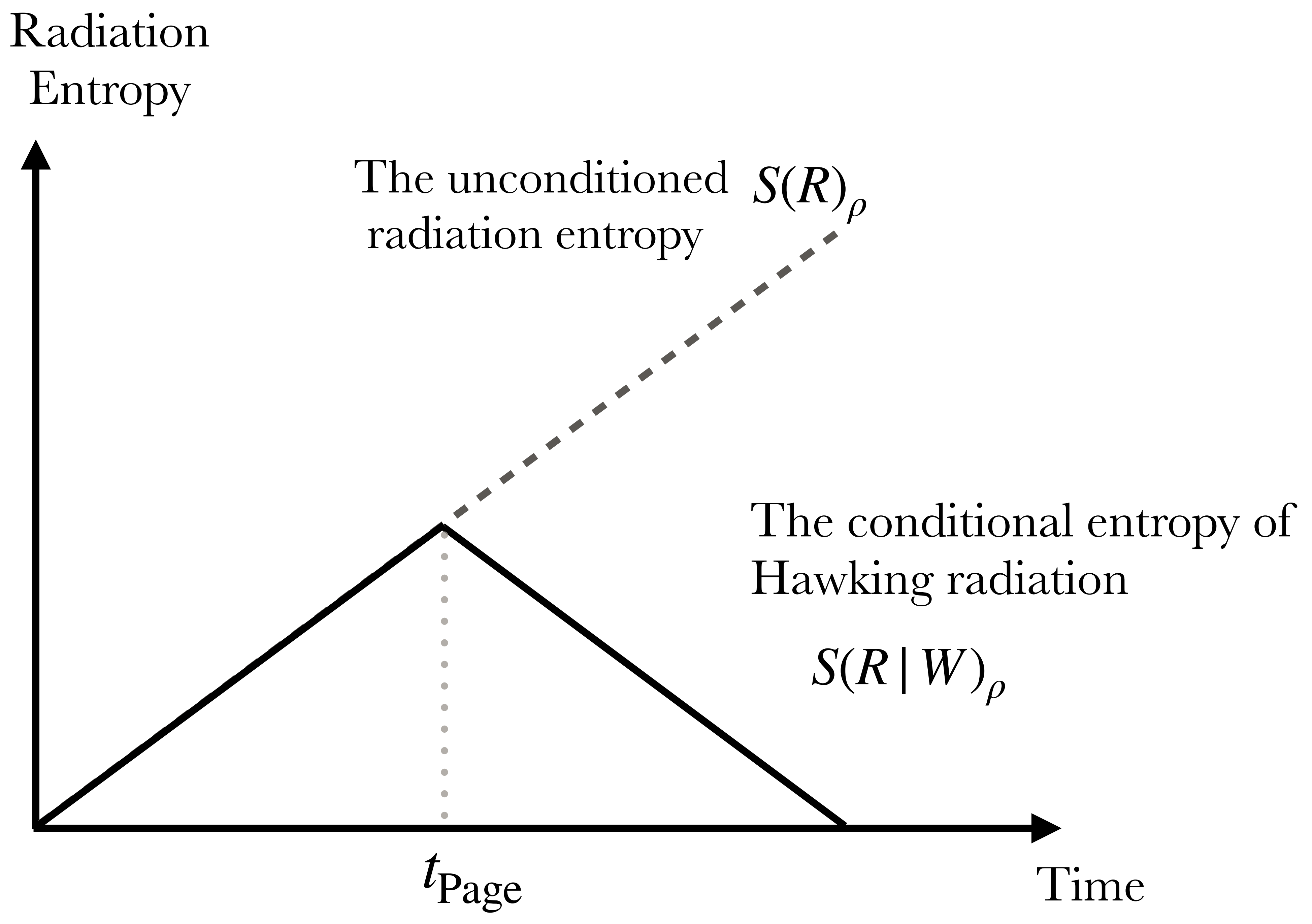
We can extend the state by introducing an **abstract reference** system W for the ensemble $\mathscr{W}$, $\rho_{R^n W} = \sum_{w \in \mathscr{W}} p_w \, \rho_{R|w}^{\otimes n} \otimes |w\rangle\langle w|_W$. [Renner, JW]

We then show that the replica trick computes the radiation entropy **conditioned** on knowing the reference W,

$$S^{\text{grav}}(R) = S(R\,|\,W)_\rho \,.$$

On the other hand, Hawking computed the entropy of the radiation **without conditioning** $S(R)_\rho$, with the mixed state $\rho_R := \sum_{w \in \mathscr{W}} p_w \rho_{R|w}$.

Both curves are correct as they refer to distinct entropies.



Viewpoint 2: the regularized entropy of Hawking radiation

How does the semiclassical GPI know about W?

Consider the total radiation entropy of N black holes. Since the contribution due to $\mathscr{W}$ doesn't scale with N, the replica trick equivalently computes the **regularized radiation entropy** of many black holes.

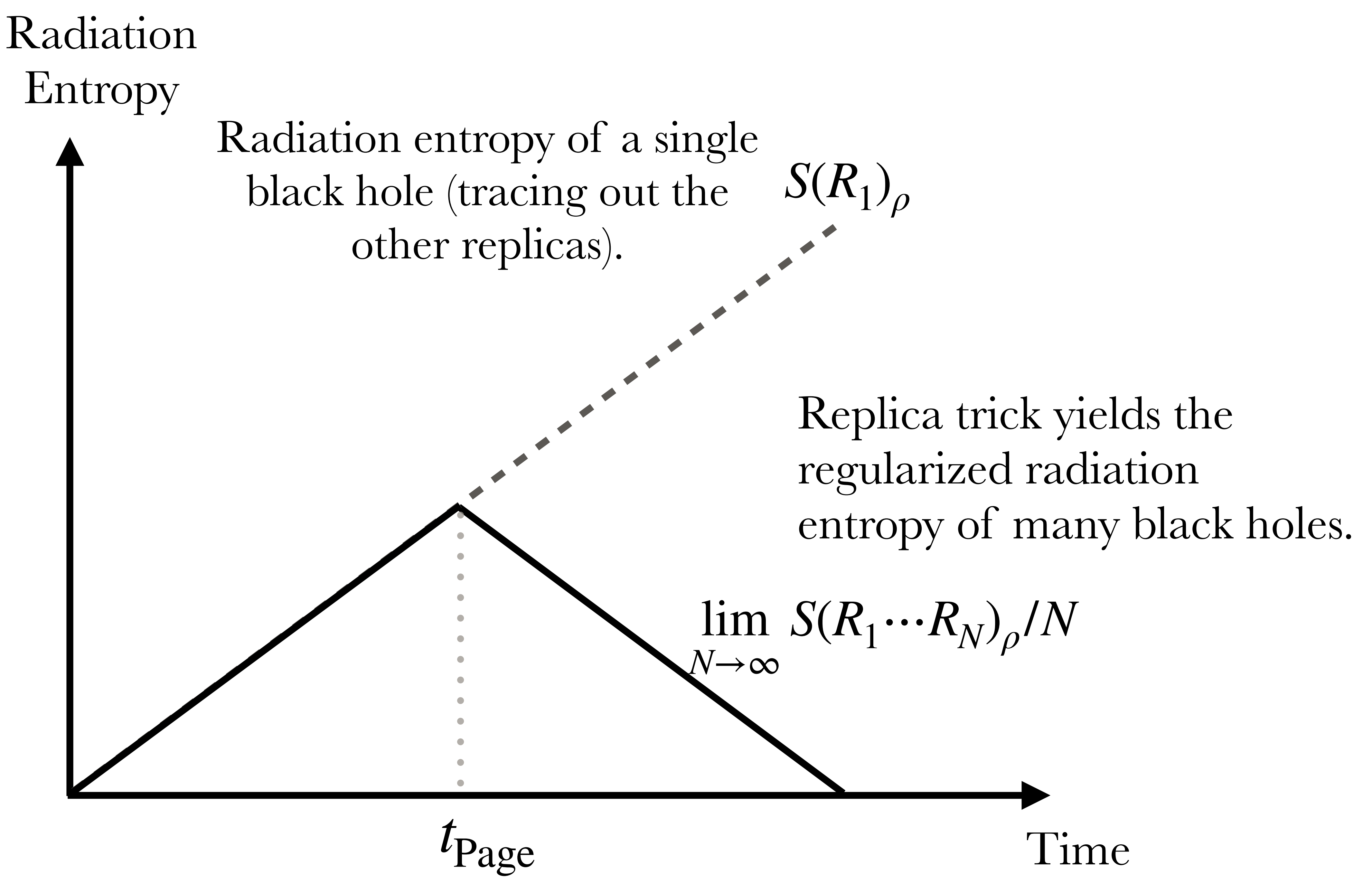
$$S^{\text{grav}}(R) = S(R\,|\,W)_\rho = \lim_{N\to\infty} S(R_1\cdots R_N)/N.$$

[Renner, JW]

W is contained in the **correlation** among the radiation systems. It is only manifest when we have many copies of the black hole, analogous to the spin example. Hence, the GPI knows $S(R\,|\,W)_\rho$ as a regularised entropy even if it doesn't need to know W a priori. Geometrically, W is encoded in the replica wormholes.

Hawking computed the entropy of a single black hole $S(R)_\rho$, with the mixture $\rho_R := \sum_{w\in\mathscr{W}} p_w \rho_{R|w}$.*

*Note that this formula doesn't make sense when we only consider one black hole, because decomposing a density matrix into an ensemble of density matrices can be arbitrary. The decomposition is only sensible when we have access to many black holes.



	Replica Trick Calculation	Hawking's Calculation
Viewpoint 1	$S(R\,\vert\,W)_\rho$	$S(R)_\rho$
Viewpoint 2	$\lim_{N\to\infty} S(R_1 \cdots R_N)_\rho / N$	$S(R_1)_\rho$

Conclusion & Outlook

Using the quantum de Finetti theorem, we give **two** interpretations of the entropy computed via the replica trick:

the radiation entropy **conditioned** on knowing some abstract reference W, which is somewhat analogous to the baby-universe idea of Marolf-Maxfield;

or the regularized radiation entropy of **many** black holes. In contrast, Hawking's ever-increasing entropy concerns a single black hole, that is less operationally relevant.

The BH information puzzle is **not fully resolved** yet. A much harder question is what would an observer see when he falls into a black hole? Most likely the semiclassical spacetime picture is no longer useful.

Thank you for listening!