Title: Demystifying the replica trick calculation of the black hole radiation entropy

Speakers: Jinzhao Wang

Series: Perimeter Institute Quantum Discussions

Date: March 16, 2022 - 11:00 AM

URL: https://pirsa.org/22030034

Abstract: The Page curve describing the radiation entropy of a unitarily evaporating black hole has recently been obtained by new calculations based on the replica trick. We analyse the discrepancy between these and Hawking's original conclusions from a quantum information theory viewpoint, using in particular the quantum de Finetti theorem. The theorem implies the existence of extra information, W, which is neither part of the black hole nor the radiation, but plays the role of a reference. The entropy obtained via the replica trick can then be identified to be the entropy S(R|W) of the radiation conditioned on the reference W, whereas Hawking's original result corresponds to the non-conditional entropy S(R). The entropy S(R|W), which mathematically is an ensemble average, gains an operational meaning in an experiment with N independently prepared black holes: for large N, it equals the regularized entropy of their joint radiation, S(R_1...R_N)/N. The discrepancy between this entropy and S(R) implies that the black holes are correlated, that is geometrically captured by the replica wormholes. In total, I will give three different interpretations of the radiation entropy calculated via the replica trick. Furthermore, I will briefly discuss the implications of ensemble interpretation in light of free probability theory, which offers the tools to deal with the effect of replica symmetry breaking in a refined calculation of the radiation entropy. (Based on the joint work (https://arxiv.org/abs/2110.14653) with Renato Renner.)

Zoom Link: https://pitp.zoom.us/j/93221648666?pwd=TkwrS0pMYjlLa090WCtCYjd0Nk9RZz09

Demystifying the replica trick calculation of the black hole radiation entropy

Jinzhao Wang

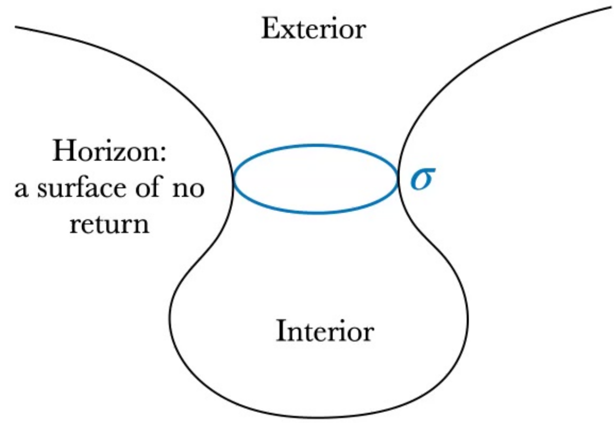
Quantum Information Seminar @ Perimeter Institute

16.03.2022

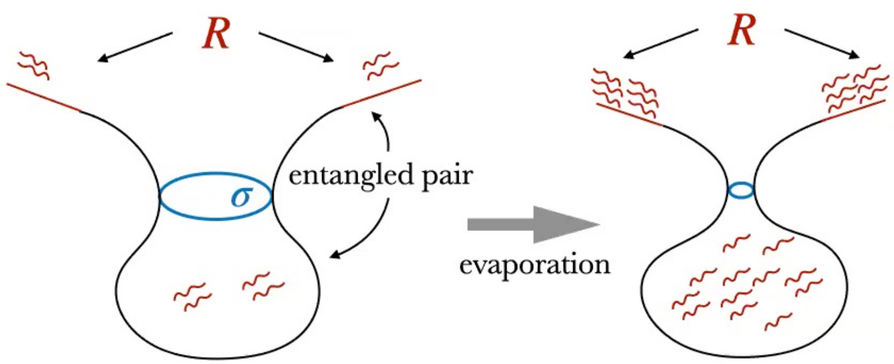
Base on the joint work with Renato Renner (arXiv: 2110.14653) and my PhD Thesis

Outline

- Review: the information puzzle, the Page curve, the replica trick and the island formula
- Some issues with the island formula
- The quantum de Finetti theorem
- Three interpretations of the radiation entropy computed via the replica trick
- Beyond the island formula



The black hole information puzzle

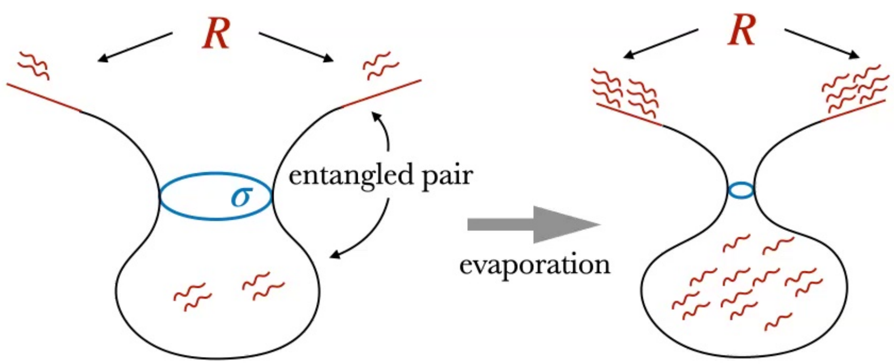


$$\rho_R \approx \rho_{\text{thermal}} \approx \frac{I}{2^k}$$

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]

The black hole information puzzle

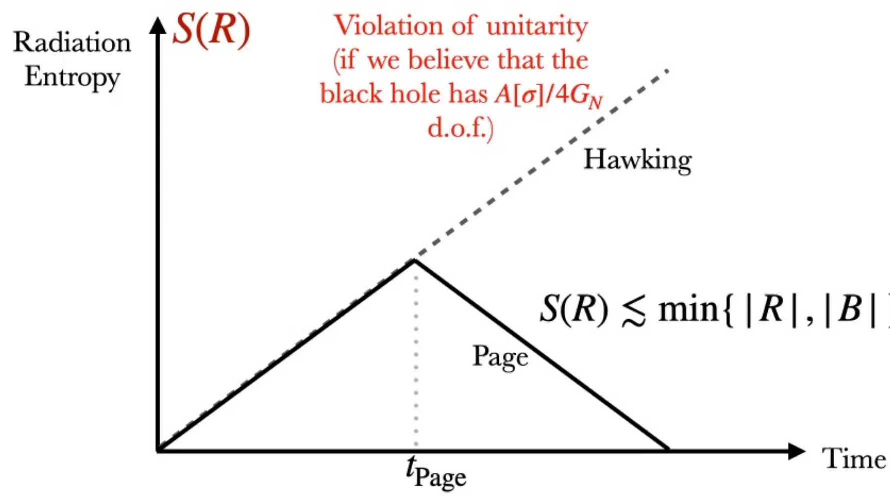


$$\rho_R \approx \rho_{\text{thermal}} \approx \frac{I}{2^k}$$

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?

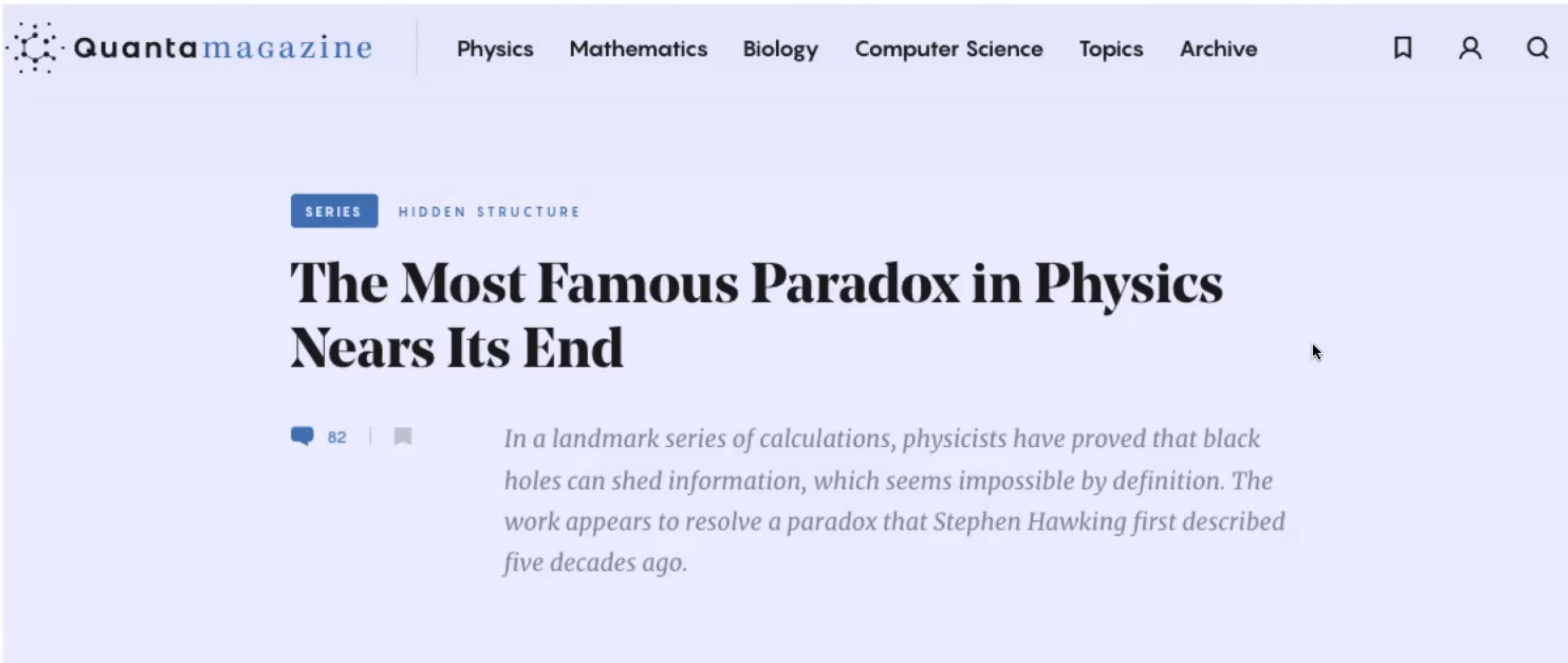


If black hole were to evolve unitarily as demanded by quantum theory, say under some typical random unitary, then the radiation entropy should follow the **Page curve**.

[Page]

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

Recent progress in calculating the Page curve


Quanta magazine

Physics Mathematics Biology Computer Science Topics Archive

SERIES HIDDEN STRUCTURE

The Most Famous Paradox in Physics Nears Its End

82

In a landmark series of calculations, physicists have proved that black holes can shed information, which seems impossible by definition. The work appears to resolve a paradox that Stephen Hawking first described five decades ago.

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

Replica trick and replica wormholes

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

replicas

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

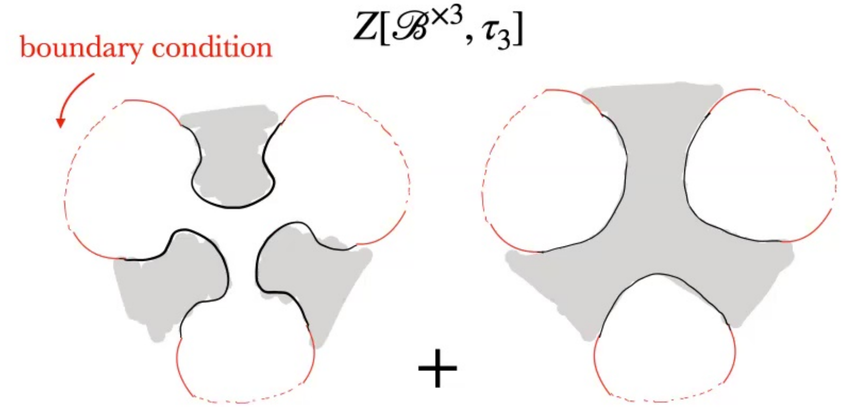
cyclic shift operator

One can use the gravitational path integral (GPI) to capture $\operatorname{Tr}\rho_R^{\otimes n}\tau_n$ 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** .

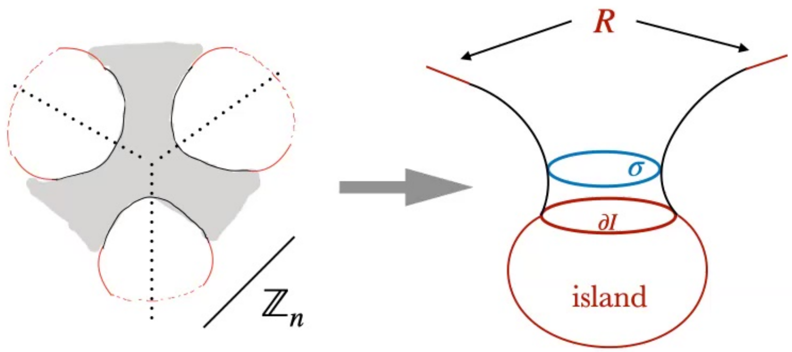
[Hawking, Gibbons, Hartle, Lewkowycz, Maldacena, Faulkner]



the Hawking saddle with disconnected black holes

the wormhole saddle with a connected replica wormhole

The Page curve from the island formula

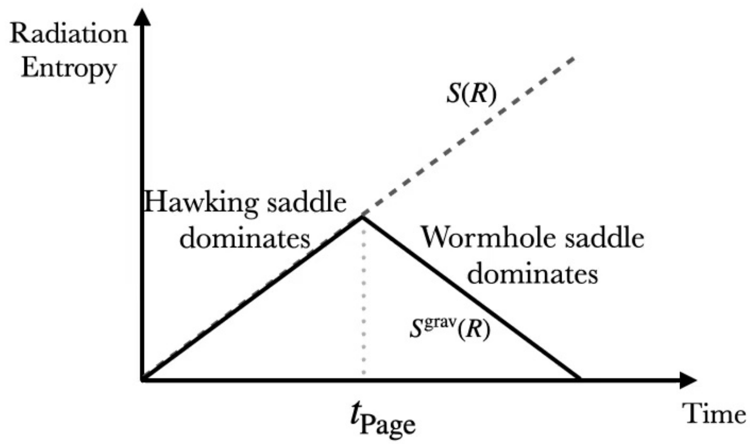


"Folding up" replica wormholes in the $n \to 1$ limit leads to the **island formula**:

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

Islands are gateways to replica wormholes

The Page curve from the island formula



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.

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

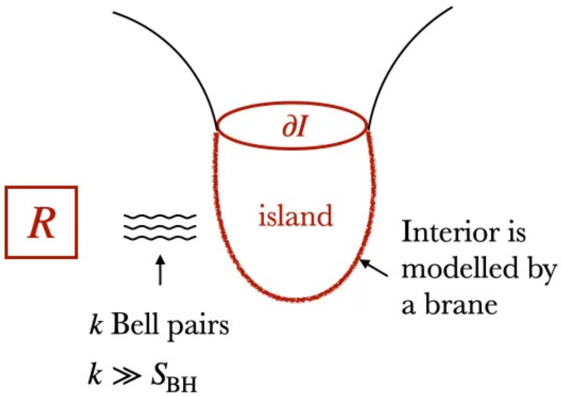
$$S^{\mathrm{grav}}(R) := \min \mathrm{ext}_I S_{\mathrm{gen}}[I] = \min \mathrm{ext}_I \left(\frac{A[\partial I]}{4G_N} + S(I \cup R) \right).$$

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,

Example: the Penington-Shenker-Stanford-Yang model

[Penington, She



For example, in the 2D black hole toy model of Penington-Shenker-Stanford-Yang (PSSY), the partition function reads

contribution from a connected geometry

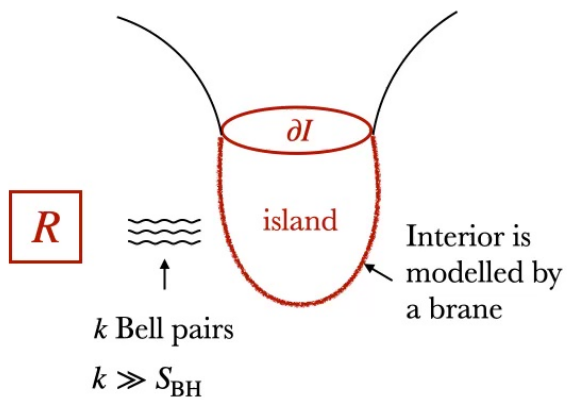
$$\frac{Z[\mathscr{B}^{\times n}, \tau_n]}{Z[\mathscr{B}^{\times n}]} = 2^k \sum_{\pi \in NC(n)} \left(\prod_{X \in \pi} \frac{\bar{Z}_{|X|}}{2^k \bar{Z}_1^{|X|}} \right) \approx \max\{2^{(1-n)k},\, 2^{(1-n)S_{\mathrm{BH}}}\}\,.$$

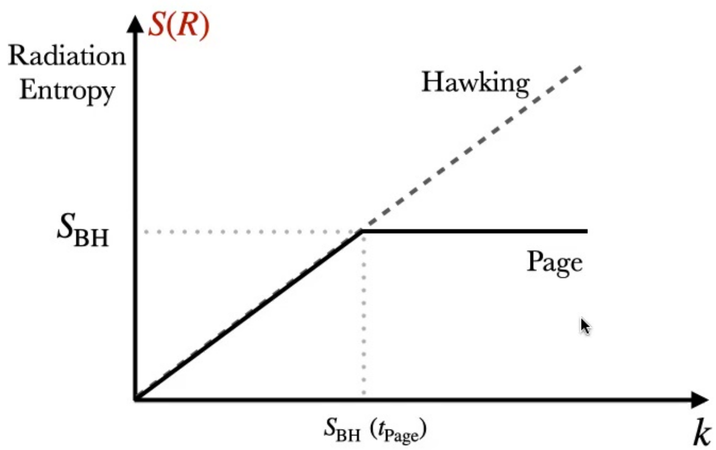
non-crossing partitions of [n], each π corresponds to one geometry compatible with the boundary conditions

either $\pi = 1$ or $\pi = \tau$ dominates

Example: the Penington-Shenker-Stanford-Yang model

[Penington, She





For example, in the 2D black hole toy model of Penington-Shenker-Stanford-Yang (PSSY), the partition function reads

contribution from a connected geometry

$$\frac{Z[\mathscr{B}^{\times n},\tau_n]}{Z[\mathscr{B}^{\times n}]} = 2^k \sum_{\pi\in NC(n)} \left(\prod_{X\in\pi} \frac{\bar{Z}_{|X|}}{2^k \bar{Z}_1^{|X|}}\right) \approx \max\{2^{(1-n)k},\, 2^{(1-n)S_{\mathrm{BH}}}\}\,.$$

non-crossing partitions of [n], each π corresponds to one geometry compatible with the boundary conditions

either $\pi = 1$ or $\pi = \tau$ dominates

Then the island formula follows,

$k \ll S_{\mathrm{BH}}$, no island,
$k \gg S_{\mathrm{BH}}$, $\partial I = \sigma$

$$S^{\mathrm{swap}}(R) = \lim_{n\to 1} \frac{1}{1-n} \log \frac{Z[\mathscr{B}^{\times n},\tau_n]}{Z[\mathscr{B}^{\times n}]} \approx \min\{k,\, S_{\mathrm{BH}}\}\,.$$

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 QI? Why is it different from Hawking's $S(R)$?

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 QI? Why is it different from Hawking's $S(R)$?

Using the quantum de Finetti theorem,
We will try to interpret $S^{\mathrm{grav}}(R)$ in three different ways to address these questions.

Black hole replicas and the Quantum de Finetti theorem

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

$$Z[\mathscr{B}^{\times n}, \tau_n] = \mathrm{Tr}\rho_{R^n}\tau_n \neq \mathrm{Tr}\rho_R^{\otimes n}\tau_n.$$

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

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

[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 $\mathscr{W}$, $\rho_{R^n W} = \sum_{w \in \mathscr{W}} p_w \rho_{R|w}^{\otimes n} \otimes |w\rangle\langle w|_W$.

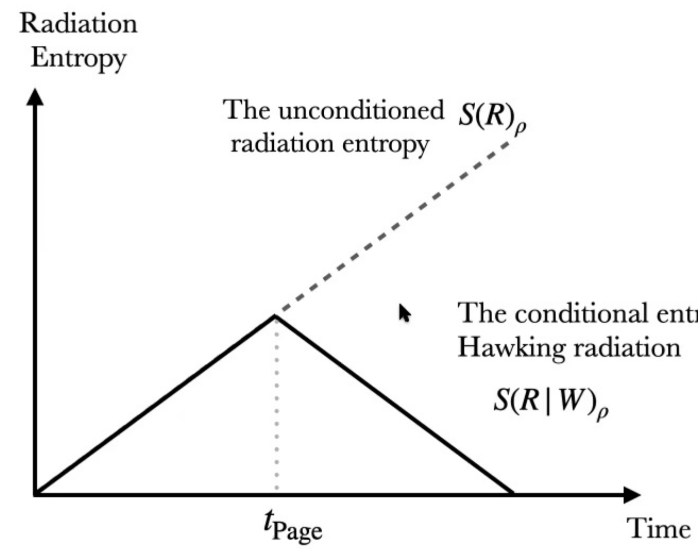
We then show that the replica trick computes the radiation entropy **conditioned** on knowing the reference W,

[Renner, JW]

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

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 quantities. However, the conditional entropy is the operationally relevant quantity if outside observers were to measure the radiation entropy. The pre measurement state has the de Finetti form as in quantum state tomography. (We cannot prepare i.i.d. copies of the mixed state ρ_R.) [Caves, Fuchs, Schack]



Viewpoint 1: The conditional entropy of Hawking radiation

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

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

[Renner, JW]

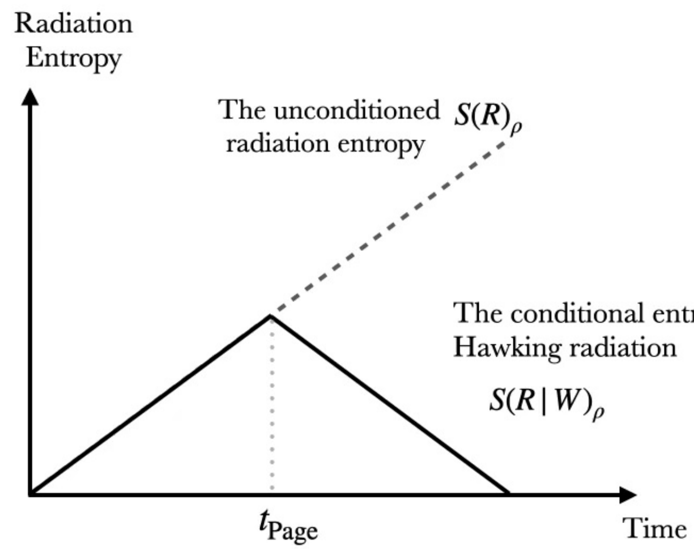
$$S^{\text{grav}}(R) = S(R \,|\, W)_\rho \,.$$

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 quantities. However, the conditional entropy is the operationally relevant quantity if outside observers were to measure the radiation entropy. The pre measurement state has the de Finetti form as in quantum state tomography. (We cannot prepare i.i.d. copies of the mixed state ρ_R.) [Caves, Fuchs, Schack]

Our result resembles the story of baby universes, that are postulated to describe the black hole interior.

[Coleman, Giddings, Strominger, Polchinski; Marolf, Maxfield]



Black hole replicas and the Quantum de Finetti theorem

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

$$Z[\mathscr{B}^{\times n}, \tau_n] = \mathrm{Tr}\rho_{R^n}\tau_n \neq \mathrm{Tr}\rho_R^{\otimes n}\tau_n.$$

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

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

[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 $\mathscr{W}$, $\rho_{R^nW} = \sum_{w \in \mathscr{W}} p_w \rho_{R|w}^{\otimes n} \otimes |w\rangle\langle w|_W$.

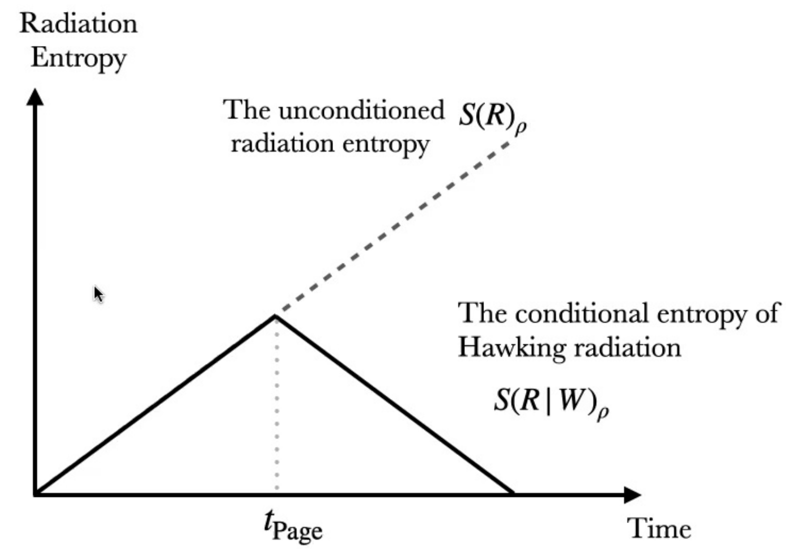
We then show that the replica trick computes the radiation entropy **conditioned** on knowing the reference W,

[Renner, JW]

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

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 quantities. However, the conditional entropy is the operationally relevant quantity if outside observers were to measure the radiation entropy. The pre measurement state has the de Finetti form as in quantum state tomography. (We cannot prepare i.i.d. copies of the mixed state ρ_R.) [Caves, Fuchs, Schack]



Our result resembles the story of baby universes, that are postulated to describe the black hole interior.

[Coleman, Giddings, Strominger, Polchinski; Marolf, Maxfield]

Viewpoint 1: The conditional entropy of Hawking radiation

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

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

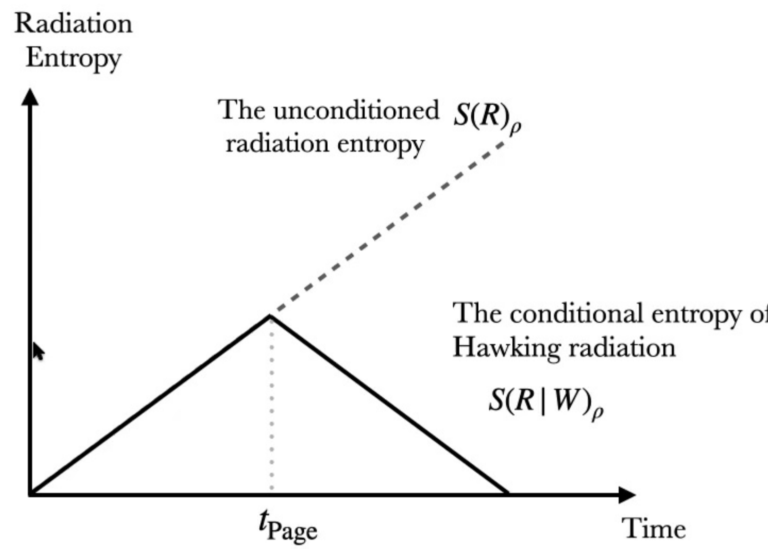
[Renner, JW]

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

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 quantities. However, the conditional entropy is the operationally relevant quantity if outside observers were to measure the radiation entropy. The pre measurement state has the de Finetti form as in quantum state tomography. (We cannot prepare i.i.d. copies of the mixed state ρ_R.)

[Caves, Fuchs, Schack]

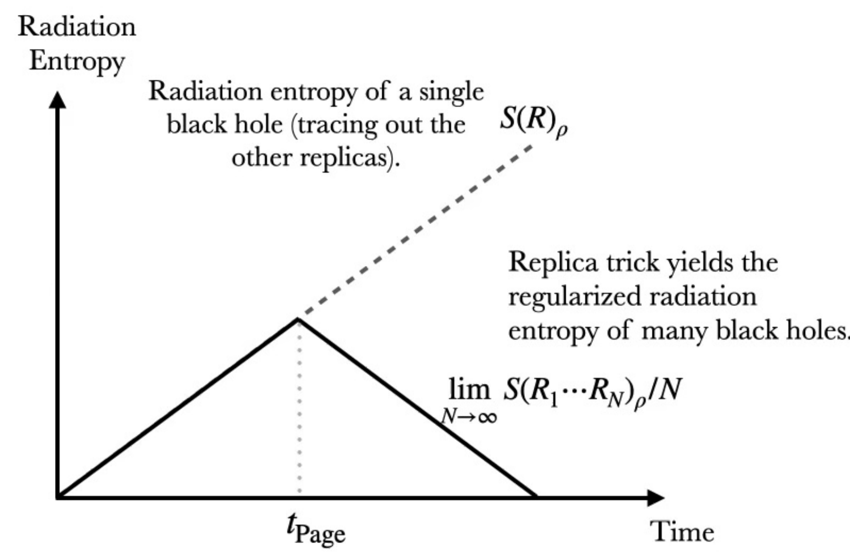


Viewpoint 2: the regularized entropy of Hawking radiation

How does the semiclassical GPI know about W?

Consider the total radiation entropy of N black holes. Since the contribution due to $\mathscr{W}$ is **bounded** (it doesn't scale with N), the replica trick equivalently computes the **regularized** radiation entropy of many black holes. [Renner, JW]

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



Viewpoint 2: the regularized entropy of Hawking radiation

How does the semiclassical GPI know about W?

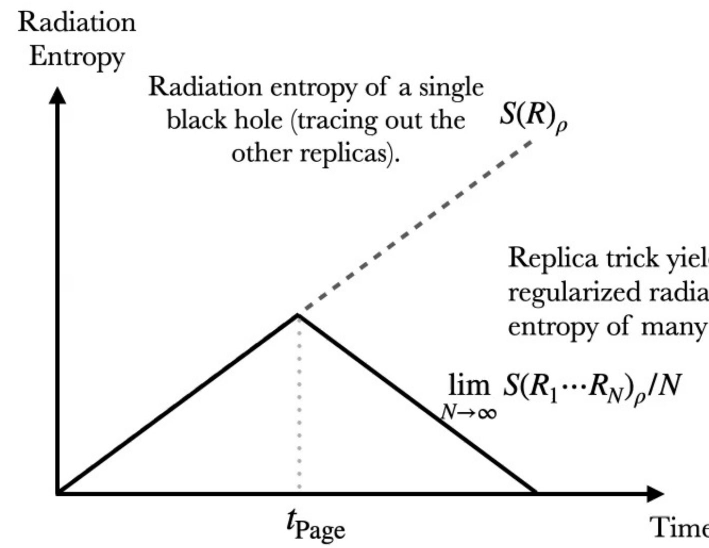
Consider the total radiation entropy of N black holes. Since the contribution due to $\mathscr{W}$ is **bounded** (it doesn't scale with N), the replica trick equivalently computes the **regularized** radiation entropy of many black holes. [Renner, JW]

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

W is contained in the **correlation** among the radiation systems. Hence, the GPI knows $S(R\,|\,W)_\rho$ even if it doesn't need to know W a priori. Geometrically, W is encoded in the replica wormholes. Note that this correlation mediated by (spacetime) wormholes are **classical**.

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.



Viewpoint 3: Radiation entropy as an ensemble average

The de Finetti state, $\rho_{R^n} = \sum_{w \in \mathscr{W}} p_w \, \rho_{R|w}^{\otimes n}$, suggests we can treat $\hat{\rho}_R$ as a **random density matrix**.

$$\hat{\rho}_R : (\mathscr{W}, \mathscr{F}, P) \to M_{N \times N}(\mathbb{C}), \quad w \mapsto \rho_{R|w} \,.$$

The replica trick concerns identical copies of the random variable $\hat{\rho}_R$. Then it follows that $S^{\mathrm{grav}}(R) = -\,\mathbb{E}[\mathrm{Tr}\hat{\rho}_R \log \hat{\rho}_R] = \mathbb{E}[S(R)_{\hat{\rho}}]$. We obtain the standard formula for the von Neumann entropy provided we treat the density matrix to be random.

The idea of **ensemble average** is considered in the literatures as a viable solution in principle. Here, it naturally follows from the de Finetti theorem. [Bousso, Tomasevic, Wildenhain]

This randomness is implicitly defined by the GPI and plausibly a general feature of quantum gravity. [Saad, Shenker, Stanford]

All we need now is the (expected) spectral measure μ_ρ to compute the expectation value

$$S^{\mathrm{grav}}(R) = -\,\mathbb{E}[\mathrm{Tr}\hat{\rho}_R \log \hat{\rho}_R] = -\int d\mu_\rho(x)\, x \log x.$$

The spectral measure can be extracted using the **free probability theory**, which describes (freely) independent non-commutative random variables, such as large random matrices with independent joint-matrix-elements distributions. It can be demonstrated in the black hole toy model of Penington-Shenker-Stanford-Yang (PSSY) that the replica trick GPI can be precisely understood as the **free moment-cumulant relation**.

Given the spectral measures, we can also compute other information-theoretic quantities such as the fidelity, relative entropies, etc in gravity, without explicitly setting the replica trick GPI every time.

	Replica Calculation	Hawking's Calculation
Viewpoint 1	$S(R\,\vert\,W)_\rho$	$S(R)_\rho$
Viewpoint 2	$\lim_{N\to\infty} S(R_1\cdots R_N)_\rho/N$	$S(R_1)_\rho$
Viewpoint 3	$\mathbb{E}S(R)_{\hat{\rho}}$	$S(R)_{\mathbb{E}\hat{\rho}}$

Beyond the island formula

Consider now a general quantum state of radiation ρ_R for a single black hole with an arbitrary spectrum. Does the island formula still apply? Is the generalized entropy still useful?

In the 2D black hole toy model of PSSY, the partition function for the replica trick can be worked out explicitly,

free cumulants of the gravitational sector $\hat{\rho}_g$ define μ_g

moments of the matter sector $\hat{\rho}_m$ define μ_m

$$\frac{Z[\mathscr{B}^{\times n}, \tau_n]}{Z[\mathscr{B}]^n} = \sum_{\pi \in NC(n)} \left(\prod_{X \in \pi} \bar{Z}_{|X|} / \bar{Z}_1^{|X|} \cdot \prod_{\bar{X} \in \bar{\pi}} \mathrm{Tr}\left(\rho_R^{|\bar{X}|}\right) \right) = \mathbb{E}[\mathrm{Tr}\hat{\rho}_R^n]$$

summing over all viable geometries labeled by non-crossing partitions π, besides the Hawking saddle ($\pi = 1$) and the wormhole saddle ($\pi = \tau$).

the Kreweras complement of π

This moment-cumulant formula defines the operation of **free multiplicative convolution** of two spectral measures corresponding to the gravitational and matter sectors respectively :

$$\mu_\rho = \mu_g \boxtimes \mu_m$$

[Voiculescu; Nica, Speicher]

The spectral measure of the radiation and thus the entropy can obtained from the free multiplicative convolution, which is unlikely to be the generalized entropy of some surface.

$$S^{\mathrm{grav}}(R) = -\int d\mu_g \boxtimes \mu_m(x)\, x \log x$$

Restoring the island formula with the one-shot entropies

When can we still trust the island formula?

$$\frac{Z[\mathscr{B}^{\times n}, \tau_n]}{Z[\mathscr{B}]^n} = \mathbb{E}[\mathrm{Tr}\hat{\rho}_R^n] = \sum_{\pi \in NC(n)} \left(\prod_{X \in \pi} \frac{\bar{Z}_{|X|}}{\bar{Z}_1^{|X|}} \cdot \prod_{\bar{X} \in \bar{\pi}} \mathrm{Tr}\left(\rho_R^{|\bar{X}|}\right) \right)$$

When one term dominates the sum,

One-shot entropies of ρ_R determine if one of the two particular saddles ($\pi = 1, \tau$) dominates over the rest.

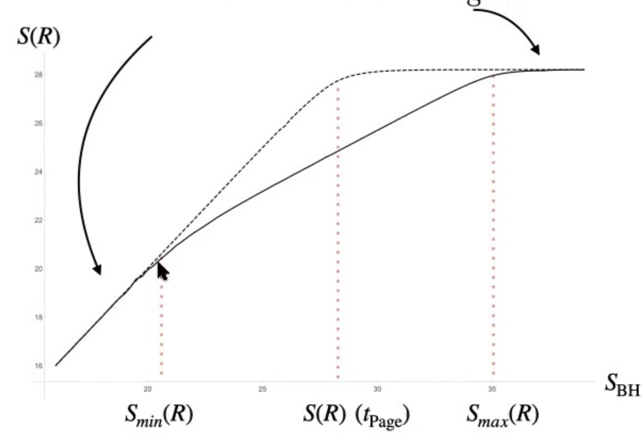
[Akers, Penington;JW]

$$\frac{Z[\mathscr{B}^{\times n}, \tau_n]}{Z[\mathscr{B}]^n} = \mathbb{E}[\mathrm{Tr}\hat{\rho}_R^n] \approx \max\{\mathrm{Tr}\rho_R^n, 2^{(1-n)S_{\mathrm{BH}}}\}$$

$$S^{\mathrm{grav}}(R) = \mathbb{E}[S(R)_{\hat{\rho}}] \approx \min\{k, S_{\mathrm{BH}}\}$$

e.g. a Page curve in the PSSY model:

The island formula holds at the regimes



The Page time generally does not indicate the relevant transition in radiation entropy.

The island formula only holds when the free multiplicative convolution **factorizes**.

Conclusion

The fine-grained entropy of the Hawking radiation is captured by the **island formula**, which evaluates the generalized entropy on the island. It yields the Page curve indicating unitary black hole evolution.

Using the de Finetti theorem, we give **three** interpretations of this fine-grained entropy: the radiation entropy **conditioned** on knowing some abstract reference W; the expectation value of the entropy of the radiation depicted by a **random** density matrix; the regularized radiation entropy of **many** black holes.

In contrast, Hawking's ever-increasing entropy concerns a single black hole, that is not operationally relevant for the outside observer collecting Hawking radiation.

Under the ensemble interpretation, we can go beyond the island formula and compute the radiation entropy more accurately using free probabilistic tools (at least in the PSSY model).

Thank you for listening!