

# Beyond islands: A free probabilistic approach

**Working Paper****Author(s):**

Wang, Jinzhao

**Publication date:**

2022-12-23

**Permanent link:**

<https://doi.org/10.3929/ethz-b-000593308>

**Rights / license:**

[Creative Commons Attribution 4.0 International](#)

**Originally published in:**

arXiv, <https://doi.org/10.48550/ARXIV.2209.10546>

# Beyond islands: A free probabilistic approach

---

**Jinzhao Wang**

*Institute for Theoretical Physics, ETH Zürich, 8093 Zürich, Switzerland*

*Stanford Institute for Theoretical Physics, Stanford University, Stanford, CA 94305*

*E-mail: [jinzhao@stanford.edu](mailto:jinzhao@stanford.edu)*

**ABSTRACT:** We give a free probabilistic proposal to compute the fine-grained radiation entropy for an arbitrary bulk radiation state, in the context of the Penington-Shenker-Stanford-Yang (PSSY) model where the gravitational path integral can be implemented with full control. We observe that the replica trick gravitational path integral is combinatorially matching the free multiplicative convolution between the spectra of the gravitational sector and the matter sector respectively. The convolution formula computes the radiation entropy accurately even in cases when the island formula fails to apply. It also helps to justify this gravitational replica trick as a soluble Hausdorff moment problem. We then work out how the free convolution formula can be evaluated using free harmonic analysis, which also gives a new free probabilistic treatment of resolving the separable sample covariance matrix spectrum.

The free convolution formula suggests that the quantum information encoded in competing quantum extremal surfaces can be modelled as free random variables in a finite von Neumann algebra. Using the close tie between free probability and random matrix theory, we show that the PSSY model can be described as a random matrix model that is essentially a generalization of Page's model. It is then manifest that the island formula is only applicable when the convolution factorizes in regimes characterized by the one-shot entropies. We further show that the convolution formula can be reorganized to a generalized entropy formula in terms of the relative entropy.

---

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
1.1	Problem: The PSSY model with non-flat entanglement spectrum	3
1.2	Solution: The free probabilistic toolkit	7
<b>2</b>	<b>A free probability primer</b>	<b>11</b>
2.1	Non-commutative probability spaces	11
2.2	Non-commutative distributions, density operators and entropies	13
2.3	Freeness	16
2.4	Free harmonic analysis	19
2.4.1	The Cauchy transform	19
2.4.2	Free additive convolution and $R$ -transform.	20
2.4.3	Free multiplicative convolution and $S$ -transform.	21
2.4.4	Free compression	22
2.5	Circular, semi-circular, quarter-circular and free Poisson distributions	24
<b>3</b>	<b>Results</b>	<b>27</b>
3.1	Replica trick as a free multiplicative convolution	28
3.2	Evaluating the free multiplicative convolution formula	32
3.3	A random matrix perspective	34
3.3.1	A model in free random variables	34
3.3.2	A random matrix model	35
3.3.3	A generalized Page's model	37
3.4	Hawking's surprisal	39
3.5	Sanity checks	42
3.5.1	The island formula and the one-shot entropies	43
3.5.2	The result of PSSY	45
3.5.3	The result of Cheng et al on random tensor networks	47
<b>4</b>	<b>Takeaway: Freeness is good</b>	<b>48</b>
<b>A</b>	<b>Additivity violation of the minimum channel output entropy</b>	<b>49</b>
<b>B</b>	<b>More on random matrices</b>	<b>52</b>
B.1	Preliminaries	52
B.2	Making contact with free probability	53
B.3	Sample covariance matrix and separable sample covariance matrix	56

---

# 1 Introduction

Recent advances in the black hole information puzzle [1, 2] feature a quantitative characterisation of the entropy of Hawking radiation that is consistent with unitarity [3–8]. The radiation entropy is shown to follow the Page curve and it resolves the *entropic information puzzle*,<sup>1</sup> which we refer to as the tension between an ever-increasing radiation entropy computed by Hawking, and the entropy following the Page curve which stops increasing after the Page time as demanded by unitarity [10, 11]. The entropy is computed using the gravitational path integral (GPI) implementing the *replica trick* [12–14]. In semiclassical gravity, the GPI can be evaluated under the saddle-point approximation, and the key insight is that a wormhole saddle connecting the replicas is generically dominant for an old evaporating black hole. The resulting entropy of Hawking radiation is described by the island formula,

$$S(\tilde{R}) = \min \text{ext}_I \left( \frac{\text{Area}[\partial I]}{4G_N} + S(I \cup R) \right). \quad (1.1)$$

This formula equates the radiation entropy to the generalized entropy evaluated at the island,  $\partial I$ , in the black hole spacetime. The surface that bounds the island is identified via an extremisation procedure and is known as the *quantum extremal surface* (QES) [15]. In particular, it identifies an island  $I$  inside the black hole after the Page time, and this allows the resulting radiation entropy  $S(\tilde{R})$  to stop increasing and hence resolving the paradox.

Some fineprints are needed to interpret this formula. The entropy that appears on the right is the entropy in the semiclassical theory. It is the von Neumann entropy of the quantum state reduced on the joint region  $I \cup R$  and we shall refer it as the state in the *effective description*, or simply as *the bulk state* if we borrow the language of holographic duality. Then the island formula is believed to capture the radiation entropy, as would be obtained from the von Neumann entropy of a density operator computed directly in a complete theory of quantum gravity. We shall refer to such a density operator as the *fundamental description* of the radiation, or as *the boundary state* again using the analogy of AdS/CFT. We add a tilde (e.g.  $\tilde{R}$ ) in the fundamental description to help distinguish it from the effective description. In general, we cannot deduce the fundamental description in full from the replica trick calculation and the best we can hope for is to extract the spectrum of its density operator.

However, the island formula is not guaranteed to hold in all circumstances. If one examines carefully its derivation [6, 7, 13, 14], an important step is to assume that the replica-symmetric (i.e.  $\mathbb{Z}_n$ -symmetric) saddles always dominate over the non-symmetric ones, such as the fully disconnected Hawking saddle before the Page time and the fully connected replica wormhole saddle after the Page time. When the non-replica-symmetric saddles aggregate a significant contribution comparable to the contributions from the replica-symmetric saddles, this assumption becomes invalid and *replica symmetry breaking* kicks in.

This is a fair simplification as far as demonstrating the qualitative features of the Page curve is concerned. However, it is recently argued and shown that the island formula doesn't compute the radiation entropy accurately when the effect of replica symmetry breaking is significant, which could occur in presence of a non-flat entanglement spectrum between the radiation and the black hole [16, 17]. In such cases, the quantum extremal surfaces and islands cannot be identified. This is well-expected because when the GPI is not dominated by a single geometry, so the entropy cannot be evaluated using a single semiclassical geometry when a superposition of distinct geometries is relevant.

---

<sup>1</sup>This does not concern other aspects of the information puzzle, such as the black hole microstates and the typical state firewall problem [9]. cf. discussions in [8].

The investigation of replica symmetry breaking in the context of black hole information problem is largely unexplored, mostly because the non-replica-symmetric saddles cannot be analysed with good control in the GPI. One exception is the toy model due to Penington-Shenker-Stanford-Yang (PSSY) [6]. It consists of a black hole in JT gravity coupled with a static end of world (EOW) brane, that carries internal degrees of freedom, modelling the black hole interior. It is perhaps the simplest setup which mimics an evaporating black hole such that the entropic information puzzle is still manifest. The upshot of the simplification is that one can make sense of summing over all saddles with good control. It is thus possible to carry out detailed quantitative analysis probing replica-symmetry, which could be challenging in other models.

### 1.1 Problem: The PSSY model with non-flat entanglement spectrum

Our goal in this work is to explicitly compute in the PSSY model the spectrum of the radiation state in the fundamental description, given *any bulk entanglement spectrum*, and thus the von Neumann or Rényi entropies of the radiation. Non-flat spectrum is relevant when we apply quantum information processing to the radiation or when the reservoir is ruled by some interesting Hamiltonian. The bottom line is that it is a more general setting to study the entropic information puzzle that was not thoroughly investigated. We shall see that this slight generalization offers some insights on the replica trick and the information puzzle.

We start by briefly reviewing the PSSY model. The black hole is described in JT gravity, which is a 2D theory of gravity that couples to a dilaton field  $\phi$ . Concretely, we work with the Euclidean JT action [6, 7, 18],

$$I_{\text{JT}}[g, \phi] = -\frac{S_0}{2\pi} \left( \frac{1}{2} \int_{\mathcal{M}} \sqrt{g} R + \int_{\partial\mathcal{M}} \sqrt{h} K \right) - \left( \frac{1}{2} \int_{\mathcal{M}} \sqrt{g} (R+2)\phi + \int_{\partial\mathcal{M}} \sqrt{h} (K-1)\phi \right), \quad (1.2)$$

where  $S_0$  is the extremal entropy of the black hole solution, i.e., the black hole entropy in the zero temperature limit. The total action is appended by the addition term describing a static brane that sits at the end of the world,

$$I_{\text{JT+EOW}}[g, \phi] = I_{\text{JT}}[g, \phi] + \mu \int_{\text{EOW}} ds \quad (1.3)$$

where  $\mu$  denotes the tension of the EOW brane.

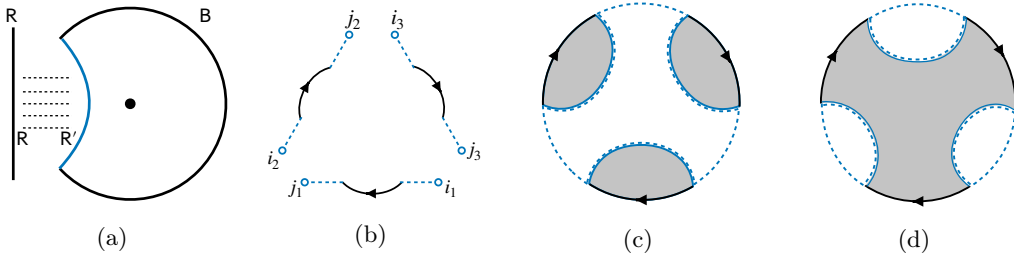
The EOW brane carries internal degrees of freedoms that are entangled with the Hawking radiation in an auxiliary  $k$ -dimensional reservoir system outside the spacetime (cf. Fig. 1a). Specifically, the scenario studied by PSSY concerns the *maximally entangled bulk state* in the effective description:

$$|\rho\rangle_{\text{R}'\text{R}} := \frac{1}{\sqrt{k}} \sum_{i=1}^k |\psi_i\rangle_{\text{R}'} |i\rangle_{\text{R}} \quad (1.4)$$

where  $|\psi_i\rangle_{\text{R}'}$  is the brane state acting as the Hawking partner that purifies the  $i^{\text{th}}$  radiation mode  $|i\rangle_{\text{R}}$ .<sup>2</sup> Here the Schmidt coefficients are chosen to be flat with the Schmidt rank  $k$ . The model does

---

<sup>2</sup>In PSSY equation (2.5), they write down the state  $|\tilde{\rho}\rangle_{\text{BR}} := \frac{1}{\sqrt{k}} \sum_{i=1}^k |\psi_i\rangle_{\text{B}} |i\rangle_{\text{R}}$  to motivate the boundary conditions for the GPI. We put a tilde to indicate that, in our terms,  $|\tilde{\rho}\rangle_{\text{BR}}$  is the state in the fundamental description, where  $|\psi_i\rangle_{\text{B}}$  is the black hole state when the brane is in state  $i$ . They have overlap of  $\mathcal{O}(e^{-S_{\text{BH}}})$  such that  $S(R)$  follows the Page curve. Here we specifically refer to the state in the effective description. It is written in Schmidt decomposition where the brane states  $|\psi_i\rangle_{\text{R}'}$  are strictly orthogonal.



**Figure 1: The PSSY model.** The setup is sketched in Fig. 1a, where the EOW brane-radiation bipartite system is labeled as R'R and the black hole-radiation system is labeled as BR. We shall consider arbitrary pure bipartite state on R'R (effective description), denoted as the dashed line, and aim to compute the entanglement spectrum between BR (fundamental description). This calculation is achieved using the replica trick that concerns for instance three replicas of boundary conditions as depicted in Fig. 1b. The EOW brane indices are contracted by the 3-swap operator  $\eta_3 = \delta_{j_1 i_2} \delta_{j_2 i_3} \delta_{j_3 i_1}$ . The dominant replica-symmetric saddles in the GPI are the fully disconnected (Hawking) saddle (Fig. 1c) and the fully connected wormhole saddle (Fig. 1d).

not describe a dynamical evaporation process. With the entanglement between the radiation and the black hole put in by hand, the PSSY model should be viewed as a snapshot of a physical black hole that is evaporating quasi-statically. To see the evolution of the radiation entropy and the Page curve, one simply tune up  $k$ . Regardless of its simplicity, the entropic information puzzle is still manifest as the radiation entropy seems to increase without bound as we tune up the bulk entanglement.

The island formula in the PSSY model is simply,

$$S(\tilde{R}) = \min\{S_{\text{BH}}, \log k\} . \quad (1.5)$$

where  $S_{\text{BH}}$  is the Bekenstein-Hawking (BH) entropy.<sup>3</sup>

According to the formula, the Page curve has two regimes with  $\log k < S_{\text{BH}}$  and  $\log k > S_{\text{BH}}$ . In the former case, the QES surface is the empty surface so we have no area term but only a bulk entropy term with contribution  $\log k$ ; and in the latter case, the QES surface sits at the horizon and we have no bulk entropy contribution  $S(I \cup R) = 0$  but only a constant area term  $S_{\text{BH}}$ . The island formula therefore provides us a sketch of the Page curve for an eternal black hole as we tune up  $k$ , which increases with  $k$  until it flattens out for  $\log k > S_{\text{BH}}$ . It's very accurate in the microcanonical ensemble and subject to subleading  $1/\beta$  correction near the Page transition in the canonical ensemble of inverse temperature  $\beta$ .

To see how this formula comes from, we need use the replica trick. The idea is to compute the moments,  $\text{Tr} \tilde{\rho}_{\text{R}}^n$ , of a hypothetical radiation density matrix  $\tilde{\rho}_{\text{R}}$ . Since we do not have such a density matrix a priori, we compute the moments by preparing replicas of the black hole-radiation system, and evaluate the expectation value of the  $n$ -swap operator  $\langle \eta_n \rangle$  using the GPI. The rule is to integrate over all the configurations compatible with the *boundary conditions* that prepares the state and imposes the observable  $\eta_n$ . Since the full path integral is challenging, we instead sum over all saddlepoint configurations to approximate the GPI. This quantity that the GPI computes can be thought of as a partition function  $\tilde{Z}_n = \langle \eta_n \rangle$  with the replica trick boundary condition.

The boundary condition is illustrated for three replicas in Fig. 1b. It consists of three copies of a asymptotic boundary (solid line) of length  $\beta$  that prepares the JT black hole at temperature  $1/\beta$  and

<sup>3</sup>Later in Section 3.1, we will provide details of how the BH entropy can be computed in the PSSY model.

the end points are joined by the sources that create the EOW brane. They carry internal degrees of freedom indicated by the dangling dashed lines. As labeled, the indices indicate the matrix elements of the radiation density operator that we'd like to prepare. The dashed lines are connected cyclically corresponding to evaluating the expectation value  $\langle \eta_3 \rangle$ . When the EOW branes form in a saddle geometry that fills into this boundary condition, the dashed lines will extend along the branes as the internal degrees of freedom are physically residing on the branes. So they form in loops as depicted in Fig. 1c. We refer them as the *EOW brane loops*.

We work in the *planar limit* which means that we ignore topologies without crossings or higher genus. This is legit when we work in the regime  $S_{\text{BH}}, \log k \gg 1$  as the nonplanar geometries incur additional factors of order  $\mathcal{O}(e^{-S_{\text{BH}}}), \mathcal{O}(1/k)$  in the on-shell action. Such planar saddles are organized by *non-crossing* (NC) partitions. (cf. Definition 4)

From each saddle, there are two contribution to the partition function  $\tilde{Z}_n$ : One from the *gravitational sector*, denoted as the gravitational partition function  $Z_n$  from an  $n$ -connected disk, i.e. a *replica wormhole* connecting  $n$  boundaries; and the other from the *matter sector*, where a length- $n$  EOW brane loop gives a contribution of  $\text{Tr}(k\rho_{\text{R}})^n$ , which always evaluates to  $k$  for the reduced state  $\rho_{\text{R}}$  of (1.4).

A quick calculation shows that when  $\log k < S_{\text{BH}}$  ( $\log k > S_{\text{BH}}$ ), the fully disconnected (connected) dominates yielding the island formula (1.5). Note that we've ignored the non-replica-symmetric saddles for the island formula, but they can be counted in and it would result in a correction of order  $1/\sqrt{\beta}$  near the transition. Nonetheless, the leading order behaviour is captured by the island formula.

Consider now a pure bulk state with an arbitrary entanglement spectrum<sup>4</sup>

$$|\rho\rangle_{\text{R}'\text{R}} := \sum_{i=1}^k \sqrt{\lambda_i} |\psi_i\rangle_{\text{R}'} |i\rangle_{\text{R}} . \quad (1.6)$$

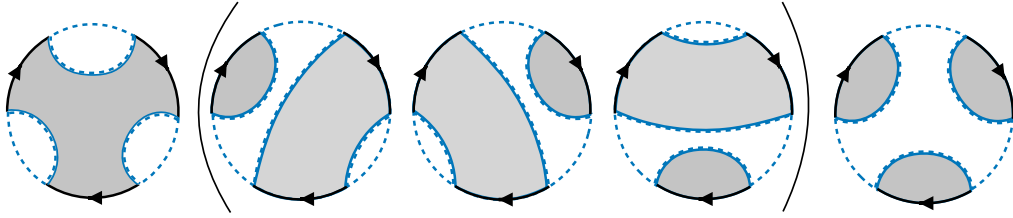
Unlike in the original PSSY with the bulk state (1.4), the EOW brane loops are now weighted by  $c_i$  and an  $n$ -connected loop evaluates to  $\text{Tr}(k\rho_{\text{R}})^n$ . We keep using  $k$  to denote the *Schmidt rank* of the bipartite state. It's shown explicitly in [17] that a bulk state non-trivial entanglement spectrum, specifically a state in superposition of two branches,<sup>5</sup> does lead to a leading order correction to the island formula.

This feature was first pointed out by Akers-Penington (AP) in the QES prescription in AdS/CFT [16]. Heuristically, the non-uniform correlation in the matter sector enhances the effect of replica symmetry breaking. The non-flatness of the spectrum is gauged by the difference between the min-entropy and the max-entropy of the bulk state, which are the entropic measures invented to characterize information-processing tasks in the one-shot regime [19].

We expect the failure of the island formula at the leading order if their difference is comparable to the area term (BH entropy). This deviation occurs in a large regime “near” the transition governed by the gap between the min-entropy and the max-entropy. In fact, the Page time, usually defined as the time when the bulk entanglement entropy is equal to the BH entropy, is not longer a relevant scale

<sup>4</sup>For a generic mixed state, which can always be considered as a marginal of a purification, we have a tripartite correlated system among the black hole, radiation and the reference. The reason we restrict to pure bulk states is that we mostly care about the quantum correlation between the radiation and the black hole in the context of the information puzzle. Also, the Page curve concerns the entanglement entropy, and this is only an operationally meaningful measure of the entanglement between the radiation and the black hole.

<sup>5</sup>This state has the Schmidt coefficients in (1.6) chosen as  $\{\lambda_i\}_{i=1}^m = x, \{\lambda_i\}_{i=m+1}^k = y$  and  $x \gg y, k \gg m$ .



**Figure 2: Saddles for three replicas.** The figure shows all the relevant five saddles for three replicas. The middle three in the bracket are the non-replica-symmetric ones which are ignored in the island formula derivation but we include in our calculation. The partition function evaluates to  $\tilde{Z}_3 = Z_3 \cdot 1^3 + 3Z_2Z_1 \cdot k^2 \text{Tr} \rho_R^2 + Z_1^3 \cdot k^3 \text{Tr} \rho_R^3$ .

to characterize the transition. It is superseded by the one-shot versions indicating when the island formula ceases to apply.

We shall now attack the problem for an arbitrary bulk entanglement spectrum in (1.6), and we need a replacement for the island formula. Implementing the replica trick gives the partition functions  $\{\tilde{Z}_n\}_{n \in \mathbb{N}}$  which admit the following form,

$$\tilde{Z}_n = \sum_{\pi \in \text{NC}_n} \left( \prod_{V \in \pi} Z_{|V|} \cdot (\text{matter sector}) \right). \quad (1.7)$$

where the contributions are organized by the non-crossing partitions. Each saddle corresponds to a configuration of wormholes that corresponds to a NC partition  $\pi$  (cf. Fig. 3), and the total contribution from the gravitational sector is simply the multiplication of the contributions  $\{Z_{|V|}\}_{V \in \pi}$  for each  $|V|$ -connected wormhole. See Fig. 2 for the example of  $\tilde{Z}_3$ .

To sort out contributions from the matter sector, we observe that the EOW brane loops define another set of cycles/partitions corresponding to a unique NC partition  $\bar{\pi} \in \text{NC}_n$  dual to each  $\pi \in \text{NC}_n$ . This is called the *Kreweras complement* of  $\pi$ , defined as follows. The rule is to replicate another set of  $[n]$  (say in red) and interlace them with the original set  $[n]$  (say in black); and then link up as many red dots as possible without crossing the black dots. It is evident that this procedure defines a unique NC partition  $\bar{\pi}$ . (See an illustration in Fig. 3b and a precise definition is provided in Section 2.3).

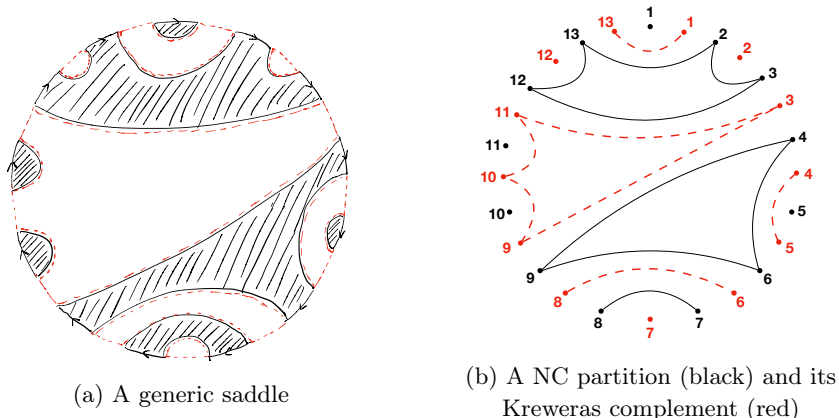
We therefore have the matter contributions also organized,

$$\tilde{Z}_n = \sum_{\pi \in \text{NC}_n} \left( \prod_{V \in \pi} Z_{|V|} \cdot \prod_{\bar{V} \in \bar{\pi}} \text{Tr} \left[ (k \rho_R)^{|\bar{V}|} \right] \right). \quad (1.8)$$

With the notion of Kreweras complement defined, this formula can be readily read off from the geometric diagrams of the saddles.

The main *hypothesis* of the replica trick is that the  $\tilde{Z}_n$ 's (with appropriate normalization) are the moments of a density operator of the radiation in the fundamental description, i.e.  $\text{Tr} \rho_R^n$ . The replica trick GPI acts like an oracle, that takes as inputs the moments of the matter density operator and the moments of the black hole thermal spectrum, and then outputs a collection of moments. Mathematically, the replica trick should be understood as a *moment problem*. Generally, there is no guarantee that this moment problem is well-posed, in the sense that the solution exists and it is unique. Here we have a rather explicit formula for its moments, we can in principle address this issue





(a) A generic saddle

(b) A NC partition (black) and its Kreweras complement (red)

**Figure 3: The combinatorics of the saddles.** The saddles are fully characterized by the set of NC partitions. Specifically, any wormhole configuration corresponds to a NC partition and the configuration of the EOW brane loops corresponds to the Kreweras complement, which is precisely defined via a partition diagram that underlies the saddle configuration. Here we illustrate a generic example that the wormholes are depicted by the partition (in black)

$\pi = (1)(2\ 3\ 12\ 13)(4\ 6\ 9)(5)(7\ 8)(10)(11) \in \text{NC}_{13}$ , and the EOW brane loops are depicted by its Kreweras complement (in red)  $\bar{\pi} = (1\ 13)(2)(3\ 9\ 10\ 11)(4\ 5)(6\ 8)(7)(12) \in \text{NC}_{13}$ .

and hopefully work out the spectrum of the density operator. The only trouble is that it isn't obvious what the formula is actually doing, and the key is to examine it through the lens of free probability.

## 1.2 Solution: The free probabilistic toolkit

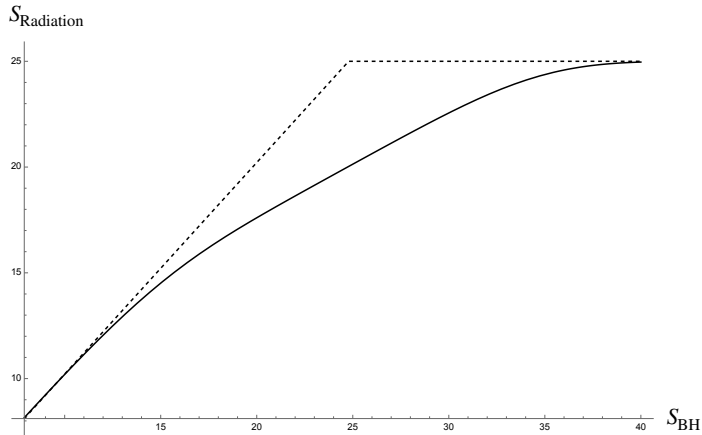
The key observation is that the combinatorial structure of how the moments are processed in (1.8) precisely matches with how the *free multiplicative convolution* between two probability distributions works (cf. (2.39)). To see this, we'd need a brief introduction of free probability theory. We shall follow Speicher's combinatorial treatment [20–22],<sup>6</sup> in which the theory is built upon the key object of *free cumulants*. We will give an introduction of non-commutative random variables, freeness and the free probabilistic toolkit in Section 2.

In order to match the GPI precisely with the free multiplicative convolution, we have to deal with the issue that the black hole thermal spectral distribution, denoted as  $\nu_b$  (cf. (3.3)), is *not* a probability distribution, because the black hole density of states in the large  $N$  limit is non-integrable and so is  $\nu_b$ . Instead, the moments of  $\nu_b$  should instead be viewed as the free cumulants of some other spectral distribution, known as the free compound Poisson distribution  $\mu_{\nu_b}$ . Note that  $\mu_{\nu_b}$  is completely determined by  $\nu_b$ , which can be extracted from the GPI in the JT gravity+EOW brane theory and is parameterized by the temperature and the brane tension. Finally, we obtain the spectral distribution of the radiation as a free multiplicative convolution, which we sometimes simply refer as the *free convolution*,

$$\nu_{\bar{r}} = \nu_r \boxtimes \mu_{\nu_b} . \tag{1.9}$$

Mathematically, this result says that the replica trick GPI implements the free multiplicative convolution for the two input (bulk) distributions effectively describing the black hole  $\nu_b$  and the

<sup>6</sup>See also [23] for a pedagogical introduction.



**Figure 4: A Page curve that largely deviates from the island formula.** The solid curve shows the values of the radiation entropy for various values of the BH entropy  $S_{\text{BH}} \approx S_0 + 4\pi^2/\beta$  with  $\beta$  fixed to be 5. The bulk entangled state is fixed to be a superposition of two branches with different Schmidt ranks,  $2^{15}$  and  $2^{35}$ , and thus its von Neumann entanglement entropy is 25. The transition doesn't occur at  $S_{\text{BH}} = 25$ , in contrast to the prediction of the island formula (1.5) shown as the dashed curve.

radiation  $\nu_r$ . The convolution formula computes the radiation entropy accurately even in cases when the island formula fails to apply. The result in turn justifies that our moment problem is indeed well-posed. We shall argue the existence and uniqueness of the resulting distribution, satisfying the premise of the replica trick. This is our main result and the details will be provided in Section 3.1.

Given (1.9), the radiation entropy simply equals to the (continuous) entropy of  $\nu_{\tilde{r}}$  (cf.(3.23)). Such a free convolution formula for the radiation entropy, unlike the island formula (1.1) or the QES formula, does not build upon the *generalized entropy* of some extremal surface  $\gamma$ ,  $S_{\text{gen}}(\gamma) = A[\gamma]/4G_N + S_{\text{bulk}}$ , such that its contribution in the partition function dominates over the rest. We learnt from our solution that generally there isn't a clean separation among the contributions so minimizing the generalized entropy over the candidate QES does not work. Instead, the partition functions associated with two competing QES, are *freely convoluted* by the GPI, so (1.9) supersedes the island formula in the PSSY model. An example of such a Page curve<sup>7</sup> is plotted in Fig. 4 to demonstrate the large corrections. The departure from the island formula agrees well with the one-shot entropy analysis. (cf. Section 3.5.1).

It seems that all we have claimed so far merely amounts to giving the replica trick (1.8) a fancy name and a fancy symbol, but really the advantage of having (1.9) is that we can use the tools from free harmonic analysis to evaluate  $\nu_{\tilde{r}}$ . This is totally analogous to how we can use Fourier transforms to evaluate convolutions between probability distributions for classical (scalar) random variables.

Generally, a free multiplicative convolution is difficult to evaluate even numerically. To do so, one needs to invoke a heavy set of machinery from the operator-valued free probability theory together with a couple of tricks. Fortunately, this particular free convolution at hand (1.9) is analytically

<sup>7</sup>Unlike the usual Page curve, we vary the black hole entropy instead of the radiation dimension  $k$  and the state of the bulk matter is fixed. Nonetheless, the same physics is captured because the Page curve in the PSSY model is only supposed to demonstrate the radiation entropy values at the snapshots of various configurations.

tractable using the standard tools from free harmonic analysis. In Section 3.2 we show in detail how it can be done. Our result is summarized as an algorithm below.

For any  $z$  in the upper complex plane  $\mathbb{C}^+ := \{z \in \mathbb{C} | \text{Im}(z) > 0\}$ , solve the following equation for  $y(z)$  with the two input distributions  $\nu_r, \nu_b$  (highlighted in red) supported on the positive real numbers  $\mathbb{R}_+ := \{x \in \mathbb{R} | x \geq 0\}$ ,

$$\int_{\mathbb{R}_+} \frac{\nu_b(x) dx}{z/x - k \int_{\mathbb{R}_+} \frac{x'}{1-x'y(z)} \nu_r(x') dx'} = y(z). \quad (1.10)$$

This is a fixed-point equation for  $y(z)$  that is numerically easy to solve. Then plug the solution  $y(z)$  into the following equation

$$G_{\nu_{\bar{r}}}(z) = \frac{1}{y} \int_{\mathbb{R}_+} \frac{\nu_r(x) dx}{1 - xy(z)}, \quad (1.11)$$

which gives the Cauchy transform of  $\nu_{\bar{r}}$ , denoted as  $G_{\nu_{\bar{r}}}$ , that is analytic on  $\mathbb{C}^+$ . The spectral distribution can be extracted using the Stieltjes inversion,

$$\nu_{\bar{r}}(x) = -\frac{1}{\pi} \lim_{\epsilon \rightarrow 0} \text{Im} G_{\nu_{\bar{r}}}(x + i\epsilon). \quad (1.12)$$

The resulting distribution does not admit a closed-form expression for a generic  $\nu_{\bar{r}}$ , and this algorithm gives the most explicit description that one can hope for. In the context of random matrix theory, our solution also applies to resolving the spectrum of a class of random matrix ensemble, known as the separable sample covariance matrix. The same algorithm was obtained using the moment method in the random matrix literature [24–27]. Here we offer a novel free probabilistic treatment that could potentially have independent interest.

The free convolution formula suggests that the quantum information encoded in the gravitational sector and the matter sector, or more generally the information encoded in two candidate QES,<sup>8</sup> should be modelled as free random variables in a non-commutative probability space, which we choose to be a *finite von Neumann algebra*. Voiculescu taught us that free random variables can be further modelled by large independent random matrices [28], so we can deduce a random matrix/tensor model from the algebraic random variables for the PSSY model. We show that this random matrix model is nothing but a generalization of Page’s model, which is perhaps the very first random tensor network model used to study the information-theoretic aspects of gravity. This is discussed in Section 3.3. We provide more details on the interplay between random matrices and free probability in Appendix B.

It is legitimate to worry that the free convolution formula we had could be something very special to the PSSY model, and it is unclear if it can be generalized to realistic black holes. To address this concern, we should make an attempt to put the entropy formula in such a way that it doesn’t explicitly entail the free convolution. We show that the convolution formula can be formulated as a generalized entropy, resembling the formulation due to Wall in his proof of the generalized second law [29]. It equates, up to a state-independent constant, the radiation entropy to the relative entropy between the boundary state and the Hartle-Hawking state. In this sense, we difference in the radiation entropy precisely measures Hawking’s surprisal in learning the actual state of the radiation while holding the hypothesis that it’s a thermal state. This is discussed in Section 3.4.

---

<sup>8</sup>It is merely an artifact of the PSSY model that among the two dominate saddle configurations, one only receives the contribution from the gravitational sector and the other only from the matter sector. We think generally the two sectors are not free from each other.

We also make contact with some relevant results in the literature, and perform some consistency checks on our proposal. In particular, we reproduce the original results of PSSY, and we show that the free convolution formula reduces to the island formula in regimes governed by the one-shot entropies, as expected for the general QES prescription [16, 17]. This is discussed in Section 3.5.

Lastly in Appendix A, we discuss an observation that an old black hole encodes a channel that violates the additivity of minimum channel output entropy. The latter was a fundamental conjecture in quantum Shannon theory [30] that is equivalent to the additivity of classical capacity of quantum channels, and it was proven false using a random channel construction by Hastings [31]. Here we see that an old black hole naturally encodes such a random channel, matching a known counterexample to the conjecture [32]. Even though for an old black hole the subleading non-replica-symmetric saddles can safely be ignored as far as the entropies are concerned, the additivity violation is only visible when all the saddles are counted. Surprisingly, summing over all saddles can be important in regimes where we usually we don't expect it to be.

## Notations

We use the lower case letters such as  $r, b, c, u, \dots$  to denote the non-commutative random variables, and they belong to a non-commutative probability spaces  $(\mathcal{W}, \tau)$  which is a finite von Neumann algebra represented as bounded operators on an infinite-dimensional Hilbert space. It is equipped with a trace  $\tau$  that gives the expectation values for the random variables. Their spectral distributions are denoted as  $\nu_r, \nu_b$ , etc. The symbol  $\mu$  is reserved for any specifically defined distributions. We use the upper case letters to denote the finite dimensional (random) matrices as  $R, B, C, U, \dots$ . We denote the normalized trace as  $\text{tr} := \frac{1}{N} \text{Tr}$ . For density operators under this notation convention, we shall not assign additional symbols to the density matrices, but instead directly use the system symbol to denote the its density matrix (such as  $R := \rho_R$ , etc). To avoid confusions, we use the Serif fonts, such as  $\mathbf{B}, \mathbf{R}$ , to denote the systems. The entropies of density operators are then simply denoted as  $S(R), S_\alpha(R)$ , etc. We shall use the tilde symbols  $\tilde{R}$  and  $\tilde{B}$  to denote density matrices of the radiation and the black hole in the fundamental description (at the boundary), and we use symbols  $R$  and  $B$  in the effective description (in the bulk).

## Related works

The large correction to the QES formula was first pointed out by AP [16]. An example of a bulk state in an incoherent superposition was given in support of the argument, but the general behaviour is left as indefinite. Later, the large correction is also shown for a similar non-flat bulk state in a coherent superposition of two branches in the PSSY model. Both calculations use the method of Schwinger-Dyson equation with the resolvent, which was first used by PSSY to derive the Page curve in a canonical ensemble. However, this method does not seem to be applicable when we deal with general non-flat bulk states. This work should be viewed as a generalization of the resolvent method by leveraging the power of free probability. The basic idea that underlies this work already appears in the dissertation of the author [33]. In this paper, a further developed study is presented. Free probability has also been used in computing various distinguishability measures for random states that resemble the black hole microstates [34].

A similar result is obtained by Cheng et al for random tensor networks (RTN). They consider RTN with bonds of non-flat entanglement spectrum modelling area fluctuations [35]. In a situation

where there are two minimal cuts homologous to a boundary subregion, the spectral distribution of the boundary reduced state asymptotically equals to the free multiplicative convolution of the limiting spectral distributions of the bonds at the two cuts, which are further convoluted with a Marchenko-Pastur distribution. We will elaborate on how their result compares to ours in Section 3.5.3. In this work, we go one step further by showing how this particular convolution can actually be evaluated in Section 3.2.

## 2 A free probability primer

Free probability theory is a probability theory of non-commutative random variables equipped with a special notion of independence called *freeness*. Before going into freeness, we first need to introduce some basics about non-commutative probability theory. It is usually formulated abstractly in terms of the algebra of random variables and their expectations. The abstraction allows the framework to accommodate classical probability theory, quantum theory and free probability theory. Here we give a minimal introduction of the relevant notions to set some backgrounds. We shall approach the subject following the treatments in [21, 22, 36]. Experts should feel free to skip over this section.

### 2.1 Non-commutative probability spaces

The most general non-commutative probability space is defined on a unital (i.e. it contains the multiplicative identity 1)  $*$ -algebra  $\mathcal{A}$ , which is algebra  $\mathcal{A}$  over  $\mathbb{C}$  endowed with an involution  $*$  such that for all  $a, b \in \mathcal{A}$  and  $z \in \mathbb{C}$ ,

$$(za)^* = \bar{z}a^*, (a+b)^* = a^* + b^*, (ab)^* = b^*a^*, (a^*)^* = a, 1^* = 1, \quad (2.1)$$

and elements satisfying  $a = a^*$  are called *self-adjoint*.

Examples of  $*$ -algebra that are particularly relevant to us are complex  $N \times N$  matrices with the involution being the conjugate transpose; and bounded complex-valued random variables over some probability space  $(\Omega, \mathcal{F}, P)$ , denoted as  $L^\infty(\Omega, \mathcal{F}, P)$ , with the involution being the complex conjugate.

For a  $*$ -algebra to be a probability space, we need to add an expectation that is mathematically a unital  $*$ -linear (i.e.  $\varphi(a^*) = \overline{\varphi(a)}$ ) function  $\varphi : \mathcal{A} \rightarrow \mathbb{C}$ .

**Definition 1.** A non-commutative probability space<sup>9</sup>  $(\mathcal{A}, \varphi)$  is defined as a unital  $*$ -algebra  $\mathcal{A}$  equipped with a unital  $*$ -linear functional  $\varphi$  that is positive and faithful.<sup>10</sup>

We call such a linear functional  $\varphi$  a *state* on  $\mathcal{A}$ . In probabilistic terms,  $\mathcal{A}$  is an algebra of non-commutative random variables, and the state gives their expectation values. Technically, what we defined is called a  *$*$ -probability space*.

A familiar example of a non-commutative probability space is quantum theory  $(\mathcal{A}, \varphi)$ , where  $\mathcal{A} \subset \mathcal{B}(\mathcal{H})$  is the operator algebra of observables as a subalgebra of bounded operators on a Hilbert space  $\mathcal{H}$ , and a vector in Hilbert space  $|\varphi\rangle \in \mathcal{H}$  defines the state  $\varphi(a) := \langle \varphi | a | \varphi \rangle, a \in \mathcal{A}$ . The state gives the expectation value for the observable  $a$ . In particular, we can consider a positive-operator-valued-measure (POVM) set  $\{m_i\}_{i \in I}$  and then  $\psi(m_i)$  gives the Born rule probability of observing the outcome  $i$ .

This definition of probability space is flexible enough to allow us to talk about both commutative and non-commutative random variables under the same footing. For example, we can take  $L^\infty(\Omega, \mathcal{F}, P)$

<sup>9</sup>The term “non-commutative” should throughout be interpreted as “potentially” non-commutative, because it can also model commutative random variables.

<sup>10</sup>A glossary of jargons: unital:  $\varphi(1) = 1$ , positive:  $\varphi(aa^*) \geq 0$  and faithful:  $\varphi(aa^*) = 0 \iff a = 0$ .

defined above, which we henceforth abbreviate  $L^\infty(\Omega)$ , and we take the linear function to be the expectation  $\mathbb{E}[\cdot]$  induced by the measure  $P$ . Another example is the algebra of  $N \times N$  (deterministic) complex matrices,  $(M_N(\mathbb{C}), \text{tr})$ , where  $M_N(\mathbb{C})$  is the algebra of complex matrices of size  $N$  and  $\text{tr}$  is the normalized trace,  $\text{tr}(X) := \frac{1}{N} \text{Tr} X$  for any  $X \in M_N(\mathbb{C})$ .

The more interesting example is to put these together to accommodate random matrices,

$$(L^\infty(\Omega) \otimes M_N(\mathbb{C}), \mathbb{E} \otimes \text{tr}) . \quad (2.2)$$

It is the algebra of  $N \times N$  random matrices with individual matrix entry being a random variable on  $(\Omega, \mathcal{F}, P)$ , and it is equipped with the expectation,

$$\mathbb{E} \otimes \text{tr}(a) := \mathbb{E}[\text{tr} a] \quad a \in L^\infty(\Omega) \otimes M_N(\mathbb{C}) . \quad (2.3)$$

In the above example for random matrices, the state  $\varphi = \mathbb{E} \otimes \text{tr}$  has the important property of being *tracial*. It means that  $\varphi(ab) = \varphi(ba)$ . Since our application of free probability theory concerns only the tracial case, we shall now restrict to such states. It is often desirable to have a *unique* tracial state as providing the notion of expectation. Furthermore, later we would like to be able to discuss functions of random variables, so we better work with an algebra that is completed to pave the way for functional calculus. To this end, we can complete the  $*$ -algebra to a von Neumann algebra, which is also referred as a  $W^*$ -algebra.<sup>11</sup>

A von Neumann algebra  $\mathcal{W}$  is defined as a unital  $*$ -algebra of bounded operators on a Hilbert space  $\mathcal{B}(\mathcal{H})$  that is closed in the weak operator topology.<sup>12</sup> We denote the commutant of  $\mathcal{W}$  in  $\mathcal{B}(\mathcal{H})$  as  $\mathcal{W}' := \{b \in \mathcal{B}(\mathcal{H}) | ab = ba, \forall a \in \mathcal{W}\}$ . The bicommutant theorem says that a  $*$ -algebra  $\mathcal{W}$  is a von Neumann algebra if and only if  $\mathcal{W} = (\mathcal{W}')'$ . The center of  $\mathcal{W}$  is defined as  $Z(\mathcal{W}) := \mathcal{W} \cap \mathcal{W}'$  and we say  $\mathcal{W}$  is a factor if the center is trivial, i.e.  $Z(\mathcal{W}) = \mathbb{C}1$ .

**Definition 2.** A (tracial)<sup>13</sup>  $W^*$ -probability space  $(\mathcal{W}, \tau)$  is defined as a von Neumann algebra  $\mathcal{W}$  equipped with the a unital tracial normal<sup>14</sup> state  $\tau$ .

Such a unital tracial normal state is *unique* when  $\mathcal{W}$  is a *finite* factor, which means it is either Type  $I_n$  or Type  $II_1$ .<sup>15</sup> A Type  $I_n$  factor is isomorphic to the familiar matrix algebra  $M_n(\mathbb{C})$ . A Type  $II_1$  factor only has representations as a subalgebra of bounded operators on an infinite dimensional Hilbert space  $\mathcal{H}$ . Moreover, there is no irreducible representations, that is to say that every vector in  $\mathcal{H}$  is entangled between the  $\mathcal{W}$  and its commutant  $\mathcal{W}' \in \mathcal{B}(\mathcal{H})$ . Nonetheless, trace does exist for every element in a Type  $II_1$  factor, so we can still make sense of density operators and entropies, which we shall introduce later. A Type  $II_1$  algebra is also the setting where free probability was first invented by Voiculescu to study the free group factors isomorphism problem [38, 39].

<sup>11</sup>One can also build a  $C^*$ -probability space on a  $C^*$ -algebra, see Chapter 3 in Nica-Speicher[20].

<sup>12</sup>The weak operator topology is the operator topology induced by the seminorms:  $a \mapsto (v, aw)$  for any  $v, \omega \in \mathcal{H}$ . Namely, a sequence/net of bounded operators  $a_i \in \mathcal{W}$  converges to  $a \in \mathcal{W}$  in weak operator topology if  $(v, a_i w)$  converges to  $(v, aw)$  for any  $v, \omega \in \mathcal{H}$ .

<sup>13</sup>More generally, we can allow non-tracial state  $\tau$  for a  $W^*$ -probability space. Since we shall only consider tracial situations, we often omit “tracial” in describing a  $W^*$ -probability space.

<sup>14</sup>A state  $\tau$  on  $\mathcal{W}$  is normal if we have  $\tau(\sup_\lambda a_\lambda) = \sup_\lambda \tau(a_\lambda)$  for any increasing nets  $(a_\lambda)_{\lambda \in \Lambda}$  of self-adjoint operators  $a_\lambda \in \mathcal{W}$ .

<sup>15</sup>(Hyperfinite) factors can be classified based on the properties of their projections. Here we give a brief summary. A Type I factor has a minimal projection, and is isomorphic to the algebra bounded operators on some Hilbert space  $\mathcal{B}(\mathcal{H})$ . A Type II factor has no minimal projection but has a finite projection, and it is further a Type  $II_1$  factor if all the projections are finite. Lastly, a Type III factor has no finite projections. cf. e.g. Jones [37] for detailed explanations.

Unless otherwise specified, we will by default work with a  $W^*$ -probability space (built on a finite von Neumann algebra) as the stage for free probability theory. For example, note that the random matrix algebra  $L^\infty(\Omega) \otimes M_N(\mathbb{C})$  is a direct integral of Type  $I_N$  factors with a center  $L^\infty(\Omega)$ , equipped with a trace  $\mathbb{E} \otimes \text{tr}$ , so it is indeed a  $W^*$ -probability space.

A non-commutative probability space  $(\mathcal{W}, \tau)$  can be abstractly defined without the need to concretely set up an underlying sample space and event space as we did for the random matrices. This allows to make universal statements at the abstract level, a desirable property especially convenient for studying random matrices that often manifests universal features asymptotically. Freeness, as we shall introduce later, is a good example of such an abstraction describing the universal behaviour of independently distributed random matrices.

## 2.2 Non-commutative distributions, density operators and entropies

For a classical random variable with compact support, knowing all its moments is equivalent to knowing its distribution. For instance, given the moments  $\mathbb{E}(X^n)$  of a real-valued random variable  $X$ , we can find a probability distribution  $\mu_X$  such that

$$\mathbb{E}(X^n) = \int x^n \mu_X(x) dx . \quad (2.4)$$

Similarly, for a collection of scalar random variables  $X_1, \dots, X_k$ , we have a joint probability distribution  $\mu_{X_1, \dots, X_k}$  that produces all the joint moments,

$$\mathbb{E}(X_{i_1} \cdots X_{i_n}) = \int x_{i_1} \cdots x_{i_n} \mu_{X_1, \dots, X_k}(x_{i_1}, \dots, x_{i_n}) dx_{i_1} \cdots dx_{i_n}, \quad \forall n \in \mathbb{N} \text{ and } i_1, \dots, i_n \in [k] . \quad (2.5)$$

Note that we can make sense of the probability distribution defined above in two different ways. Given an underlying Kolmogorov triple  $(\Omega, \mathcal{F}, P)$ ,  $\mu_{X_1, \dots, X_k}$  is nothing but the push forward of  $P$  under the measurable map  $\omega \in \Omega \mapsto (X_1(\omega), \dots, X_k(\omega)) \in \mathbb{R}^k$ . We call this an *analytic distribution* and this is what the distributions mean in the integral formulas above.

More abstractly, if we choose to work without referring to an underlying sample space but only with the expectations of our random variables, then the notion of probability distribution should generally be understood algebraically as the following linear map

$$\mu_{X_1, \dots, X_k} : p \in \mathbb{C}[x_1, \dots, x_k] \mapsto \mathbb{E}[p(X_1, \dots, X_k)] \in \mathbb{R} \quad (2.6)$$

where  $\mathbb{C}[x_1, \dots, x_k]$  denotes the ring of commutative polynomials in  $k$  indeterminates  $x_1, \dots, x_k$ .

We shall call this the *algebraic distribution* and this is the appropriate notion that can be generalized to the non-commutative setting, in which the Kolmogorov triple is usually abstracted away.

Given a collection of non-commutative self-adjoint random variables  $a_1, \dots, a_k \in (\mathcal{W}, \tau)$ , their *non-commutative distribution* is thus defined as the linear map

$$\mu_{a_1, \dots, a_k} : p \in \mathbb{C}\langle x_1, \dots, x_k \rangle \mapsto \tau(p(a_1, \dots, a_k)) \in \mathbb{C} \quad (2.7)$$

where  $\mathbb{C}\langle x_1, \dots, x_k \rangle$  denotes the ring of non-commutative polynomials in  $k$  indeterminates  $x_1, \dots, x_k$ . By linearity, we can equivalently define the non-commutative distribution as the collection of all their *joint moments*  $\tau(a_{i_1} \cdots a_{i_n}), \forall n \in \mathbb{N}, i_1, \dots, i_n \in [k]$ .

Generally, we do not have the analytic counterpart to a non-commutative distribution. Nonetheless, this is possible for the case of a single bounded<sup>16</sup> random variable.<sup>17</sup> We can find an analytic distribution from its moments as in the classical case. Mathematically this is known as a *Hausdorff moment problem*, which asks for a unknown compactly supported distribution that has the given moments. The solution is unique if it is solvable.<sup>18</sup>

Consider a bounded single self-adjoint random variable  $a \in (\mathcal{W}, \tau)$ . Given its moments  $\tau(a^n)$ , one can find a unique measure  $\nu_a$  with compact support on  $\mathbb{R}$  such that

$$\tau(p(a)) = \int p(x)\nu_a(x)dx \quad (2.8)$$

for any polynomial  $p \in \mathbb{C}\langle x \rangle$ . The support is explicitly given by the spectral radius of  $a$  as  $[-\rho(a), \rho(a)]$  where  $\rho(a)$  is the spectral radius of  $a$  (cf. Footnote 16).

In fact, this claim holds generally for a  $*$ -probability space.<sup>19</sup> However, the advantage of completing a  $*$ -algebra to a von Neumann algebra is that we can use these polynomials as building blocks to extend (2.8) to any continuous functions on the support of  $\nu_a$ . We can define

$$\tau(f(a)) := \int f(x)\nu_a(x)dx \quad (2.9)$$

The operator  $f(a)$  on the LHS is defined via a unital  $*$ -algebra homomorphism  $\Phi_a$  mapping from  $f \in L^\infty(\mathbb{R}, \nu_a)$  to  $f(a) \in \mathcal{W}$ .  $\Phi_a$  is known as the *functional calculus* for  $a$ . Since polynomials are dense in the continuous functions over supported over  $[-\rho(a), \rho(a)]$ , (2.9) uniquely determines  $\Phi_a$ . (cf. Theorem 3.1 in Nica-Speicher [20].) We shall refer to this analytic distribution  $\nu_a$  as the *spectral distribution* of  $a$ . Such an extension to continuous functions is necessary for us to study spectral functions like the von Neumann entropy.

We now discuss how to define entropies for states acting on a  $W^*$ -probability space. This was first studied by Segal [40]. (See also a recent article on this subject [41].) In a  $W^*$ -probability space, the *density operator*  $a \in (\mathcal{W}, \tau)$ <sup>20</sup> of a normal faithful state  $\varphi$  on  $\mathcal{W}$  is defined via the expectation  $\tau$  via  $\varphi(x) = \tau(xa)$  and has normalization  $\tau(a) = 1$ . Such assignment of the density operators is unambiguous as the expectation  $\tau$  is unique.

The von Neumann entropy of the state  $\varphi$  is then defined as usual

$$S(\varphi) = S(a) := -\tau(a \log a) := - \int x \log x \nu_a(x)dx. \quad (2.10)$$

where we've used (2.9) for  $f(x) = -x \log x$ .

<sup>16</sup>The spectral radius  $\rho(a)$  of a self-adjoint random variable in  $(\mathcal{W}, \tau)$  is defined as  $\rho(a) := \lim_{n \rightarrow \infty} |\tau(|a|^{2n})|^{1/2n}$ , and the Cauchy-Schwarz inequality implies that the moments are (exponentially) bounded by  $\rho(a)$ ,  $\tau(a^n) \leq \rho(a)^n$ . We say a self-adjoint random variable  $a$  is bounded if its spectral radius is finite.

<sup>17</sup>Self-adjointness is not necessary here for the analytical distribution to exist. We choose to focus on self-adjoint variables here because these are the most relevant for us. Generally, for any normal operator (i.e. it commutes with its involution) in a  $*$ -probability space, an analytic distribution can also be constructed from the  *$*$ -moments*,  $\tau(a^k a^{*l})$ ,  $\forall k, l \in \mathbb{N}$ . See e.g. Chapter 1 in [20].

<sup>18</sup>For a collection of moments  $\{m_n\}_{n=0}^\infty$  to define a unique probability distribution, it needs to satisfy Carleman's condition,  $\sum_{n=1}^\infty m_{2n}^{-1/2n} = \infty$ . This condition is always satisfied for a compactly supported probability distribution.

<sup>19</sup>See Theorem 2.5.8 in [36] for the more general statement.

<sup>20</sup>When the density operator  $a$  is unbounded, it does not belong to the algebra  $\mathcal{W}$  itself but is *affiliated* to the algebra, which means that bounded functions of  $a$  belong to  $\mathcal{W}$ . For simplicity, we assume that our density operators have bounded spectra.



Similarly, the  $\alpha$ -Rényi entropies are defined as

$$S_\alpha(\varphi) = S_\alpha(a) := \frac{1}{1-\alpha} \log \tau(a^\alpha) := \frac{1}{1-\alpha} \log \int x^\alpha \nu_a(x) dx, \quad \forall \alpha \in \mathbb{R}_+ . \quad (2.11)$$

The von Neumann entropy so defined take values in  $[-\infty, 0]$ , where the maximal is obtained at the tracial state  $\tau$  itself,  $S(\tau) = S(1) = 0$ . It differs from the entropy of density matrices in the usual quantum information convention. Consider now the usual convention for a finite dimensional matrix algebra  $(M_N(\mathbb{C}), \text{Tr})$ , and the density matrix  $A \in (M_N(\mathbb{C}), \text{Tr})$  of a tracial state  $\varphi$  is normalized by  $\text{Tr}A = 1$ , against the density matrix  $A_N \in (M_N(\mathbb{C}), \text{tr})$  of the same state  $\varphi$  normalized by  $\text{tr}A_N = 1$ . It follows that  $A_N = NA$  and

$$S(A) := -\text{Tr}A \log A = \log N - \tau A_N \log A_N = S(A_N) + \log N \approx S(a) + \log N . \quad (2.12)$$

Therefore, as we go to infinite dimensions  $N \rightarrow \infty$ , if we have  $A_N \rightarrow a \in (\mathcal{W}, \tau)$ , then  $S(A)$  typically diverges whereas  $S(A_N) \rightarrow S(a)$  remains finite.

We are able to compute the entropy  $S(a)$  directly from the freely convoluted spectral distribution. In order to obtain the standard entropy value  $S(A)$  for some  $N \times N$  density matrix  $A$ , we simply add back  $\log N$ . This is a good approximation if  $N$  is large enough such that  $S(A_N)$  is close to its asymptotic value  $S(a)$ .

Finally, we'd like to briefly sketch how to define the relative entropy between two faithful normal states  $\varphi$  and  $\psi$  on a von Neumann algebra  $\mathcal{W}$ . Readers can refer to the review article [42] for details.

Consider two faithful normal states  $\varphi$  and  $\psi$  on a von Neumann algebra  $\mathcal{W}$  that is not necessarily finite, they can be represented as cyclic and separating vectors  $|\varphi\rangle, |\psi\rangle \in \mathcal{H}$  in some Hilbert space  $\mathcal{H}$ ,<sup>21</sup> such that for any  $x \in \mathcal{W}$ ,

$$\varphi(x) = \langle \varphi | x | \varphi \rangle, \quad \psi(x) = \langle \psi | x | \psi \rangle . \quad (2.13)$$

Now we need to introduce the Tomita operator  $S_{\varphi|\psi}$ , defined via

$$S_{\varphi|\psi} a |\varphi\rangle = a^* |\psi\rangle, \quad \forall a \in \mathcal{W} \quad (2.14)$$

The Tomita operator admits the following polar decomposition,

$$S_{\varphi|\psi} = J_{\varphi|\psi} \Delta_{\varphi|\psi}^{\frac{1}{2}} \quad (2.15)$$

where  $J_{\varphi|\psi}$  is an anti-unitary operator that sends  $\mathcal{W}$  to  $\mathcal{W}'$ , known as the relative modular conjugation, and  $\Delta_{\varphi|\psi} = S_{\varphi|\psi}^* S_{\varphi|\psi}$  is the relative modular operator. Now we can define the relative entropy

$$D(\varphi||\psi) := -\langle \varphi | \log \Delta_{\varphi|\psi} | \varphi \rangle . \quad (2.16)$$

For a finite von Neumann algebra  $(\mathcal{W}, \tau)$ , the relative entropy can also be expressed in terms of the corresponding density operators using the more familiar Umegaki's formula,

$$D(\varphi||\psi) = D(a||b) := \tau(a \log a) - \tau(a \log b) . \quad (2.17)$$

where  $a$  and  $b$  are the density operators of  $\varphi$  and  $\psi$  respectively.

---

<sup>21</sup>Cyclic means  $\{a|\varphi|a \in \mathcal{W}\}$  is dense in  $\mathcal{H}$  and separating means  $a|\varphi\rangle = 0$  if and only if  $a = 0$ . Such a (standard) representation can be constructed using the GNS construction [43, 44].

### 2.3 Freeness

We have discussed how to make sense of the spectral distribution of a single bounded self-adjoint random variable. Can one go beyond the case of a single random variable? This calls for an appropriate notion of independence among non-commutative random variables. As we shall see, the very notion of *freeness* is an appropriate one that is alternative to the more familiar independence, and free probability theory is the study of *free* non-commutative random variables.

**Definition 3** (Freeness). *A collection of non-commutative random variables  $a_1, \dots, a_d$  in a  $W^*$ -probability space  $(\mathcal{W}, \tau)$  is free if*

$$\tau([p_1(a_{i_1}) - \tau(p_1(a_{i_1}))] \cdots [p_k(a_{i_k}) - \tau(p_k(a_{i_k}))]) = 0 \quad (2.18)$$

for any polynomials  $p_1, \dots, p_k$  and indices  $i_1, \dots, i_k \in [d]$  with no adjacent ones equal.

Let's unpack this definition of freeness. In words, it says “the alternating product of centered random variables is centered”. As advocated by Speicher, we can think of freeness as a convenient rule to allow the joint moments to be expressed in the individual moments of the individual variables. For example, the definition (2.18) implies that for free  $a$  and  $b$ ,

$$\begin{aligned} \tau(ab) &= \tau(a)\tau(b), & \tau(aba) &= \tau(a^2)\tau(b), & \tau(aabb) &= \tau(a^2)\tau(b^2) \\ \tau(abab) &= \tau(a^2)\tau(b)^2 + \tau(b^2)\tau(a)^2 - \tau(b)^2\tau(a)^2. \end{aligned} \quad (2.19)$$

This is not immediately obvious that (2.18) implies one can break down the joint moment of any word into the individual moments of its free letters, but we shall later show that this is indeed true. Since the moments define the non-commutative distributions, freeness should be understood as a notion of independence in the sense that the joint distribution of several free random variables is completely determined by the distribution of individual ones.

Note however that freeness is distinct from the notion of (tensor/Cartesian) independence, that concerns scalar random variables<sup>22</sup> or commutative objects more generally. Even for a non-commutative probability theory, such as quantum theory  $(\mathcal{A}, \varphi)$ , independence concerns commuting self-adjoint operators representing observables that are simultaneously measurable. We say two commuting observables  $a$  and  $b$  are independent in  $(\mathcal{A}, \varphi)$  if  $\varphi(a^{m_1} b^{n_1} a^{m_2} b^{n_2} \dots) = \varphi(a^{m_1+m_2+\dots})\varphi(b^{n_1+n_2+\dots})$ .<sup>23</sup> This doesn't happen say when  $\varphi$  is a Bell state shared by Alice and Bob and  $a, b$  are spacelike separated local observables of Alice and Bob. Their observations (measurement outcomes) are therefore correlated.

On the other hand, freeness (2.18) gives a distinct rule for how the joint moments should factorize for non-commutative random variables. We can use the following example to illustrate their difference. Consider two independent commuting random variables in a  $W^*$ -probability space  $(\mathcal{W}, \tau)$  with zero means but positive second moments. Then  $\tau(abab) = \tau(a^2)\tau(b^2) > 0$ , so they cannot be free because freeness demands  $\tau(abab) = \tau(a^2)\tau(b)^2 + \tau(b^2)\tau(a)^2 - \tau(b)^2\tau(a)^2 = 0$ . In fact, one can show that for a pair of commutative random variables to be free, one of them has to be a scalar, and a scalar is free from anything.<sup>24</sup> Hence, freeness and independence are almost orthogonal notions and neither is stronger than the other.

<sup>22</sup>Scalars are operators that are proportional to the identity.

<sup>23</sup>Rather than talking about independence between two observables for a given state, one can also discuss notions of independence between two observable subalgebras associated with different subsystems for any states. They are commonly formulated as the ability to independently prepare states for each subsystem. There are various ramifications of such notions and commutativity between the subalgebras are not generally demanded. See [45] and references therein.

<sup>24</sup>See for example Proposition 1.10 and 1.11 in [22].

Despite their sheer difference, freeness is a sensible counterpart of independence that is tailor-made for non-commuting random variables. This is because the key theorems and properties about independent random variables naturally have their free analogs for free random variables. Examples are various limit theorems, such as the free central limit theorem [46] and the law of large numbers [47, 48], the free convolutions of distributions [46, 49] and classification of infinitely divisible and stable laws [50, 51]. An example that is particularly relevant for us is the *free Poisson limit theorem*. It is the non-commutative analog of the Poisson limit theorem that asserts the sum of Bernoulli random variables (coin flips) asymptotically approaches a Poisson random variable. We will discuss it later in Section 2.5.

Conversely, given any probability measure  $\nu$  with compact support on  $\mathbb{R}$ , it's always possible to construct a  $W^*$ -probability space  $(\mathcal{W}, \tau)$ , such that there exists some self-adjoint  $a \in \mathcal{W}$  and  $\nu_a = \nu$ . (cf. e.g. Facts 4.2 in [22]).

Furthermore, with a collection of non-commutative probability spaces  $(\mathcal{W}_i, \tau_i)_{i \in [n]}$ , one can always amalgamate them into a larger probability space  $(\mathcal{W}, \tau)$ , such that  $\mathcal{W}_i \subset \mathcal{W}$ ,  $\tau(a_i) = \tau_i(a_i)$ ,  $\forall a_i \in \mathcal{W}_i$  and any collection of random variables drawn individually from these subalgebras,  $\{a_i\}, a_i \in \mathcal{W}_i$ , are free. (cf. e.g. Lemma 2.5.20 in Tao [36]). This is achieved by the *free product*,<sup>25</sup> denoted as  $\mathcal{W} = *_{i \in [n]} \mathcal{W}_i, \tau = *_{i \in [n]} \tau_i$ . We shall spare any details on free product here and interested readers shall consult chapters 5-7 in Nica-Speicher [20]. It allows us to introduce free random variables of specified distributions in a joint probability space, just like we are used to do in classical probability theory. The only difference is that we now replace the Cartesian/tensor product by free products.

Free probability was born in the field of operator algebra [38]. Nonetheless, it was later realized by Voiculescu that it has an intimate connection with random matrix theory. In Voiculescu's seminal work in 1991 [28], he proved the first result connecting between the asymptotic random matrices and free probability theory. (cf. Theorem 2 for the precise statement in Appendix B)

Roughly speaking, Voiculescu's theorem implies that random matrices with classically independent matrix elements tend to be free in the asymptotic limit. In fact, the asymptotic freeness would not hold without the non-commutativity as we argue above. Heuristically, we can say that any finite-dimensional random matrices are still too commutative to be free. When studying an abstract non-commutative random variable, it is often useful to use a sequence of random matrices as a model for concreteness. We will later also make contact with random matrices for a better understanding of the free probabilistic calculations. We postpone more details regarding this connection to Appendix B.

At the level of moments, it is somewhat cumbersome to describe freeness using (2.18). We now introduce an alternative characterization in terms of the *free cumulants*. We first define the notion of non-crossing (NC) partitions.<sup>26</sup>

**Definition 4.** A partition  $\pi$  of the set  $[n]$  is defined as  $\pi = \{V_i\}_i$  such that the blocks  $V_i$  satisfy  $V_i \neq \emptyset \forall i, V_i \cap V_j = \emptyset \forall i \neq j$  and  $\cup_i V_i = [n]$ . A partition  $\pi$  is non-crossing if it does not have  $a_1 < b_1 < a_2 < b_2$  for some  $a_1, a_2 \in V_i$  and  $b_1, b_2 \in V_j$  with  $i \neq j$ .

We denote the set of NC partitions of  $[n]$  as  $\text{NC}_n$ . We use  $|\pi|$  to denote the cardinality of  $\pi \in \text{NC}_n$ . Given a finite set  $[n]$  of size  $n$ , consider now doubling it with additional elements  $\bar{1}, \dots, \bar{n}$  and interlace

<sup>25</sup>Here we stick to the standard terminology. Note that the term free product is used instead for the free multiplicative convolution in Cheng et al [35], which should not be confused with the algebraic operation we are referring to here.

<sup>26</sup>The non-crossing partitions are in one-to-one correspondence with the non-crossing permutations: For each NC partition  $\pi = \{V_i\}_i$ , we let the NC permutation be defined with cycles  $(V_i)_i$  such that each cycle  $(V_i)$  consists of the elements in the block  $V_i$  arranged in an increasing order.

them with  $1, \dots, n$  in an alternating way,  $1, \bar{1}, \dots, n, \bar{n}$ . For any  $\pi \in \text{NC}_n$ , its *Kreweras complement* [52],  $\bar{\pi} \in \text{NC}(\bar{1}, \dots, \bar{n}) \cong \text{NC}_n$ , is defined as the biggest<sup>27</sup> element in  $\text{NC}(\bar{1}, \dots, \bar{n})$  such that  $\pi \cup \bar{\pi} \in \text{NC}(1, \bar{1}, \dots, n, \bar{n})$ .<sup>28</sup> See Fig. 3b for an illustration. Note that generally for any two partitions  $\pi_1, \pi_2 \in \mathcal{P}_n$ , we have  $|\pi_1| + |\pi_2| \leq |\pi_1 \pi_2| + n$ . We have the equality saturated  $|\pi| + |\bar{\pi}| = n + 1$  for and only for a NC partition and its Kreweras complement.<sup>29</sup>

We are now ready to define free cumulants.

**Definition 5.** For a unital linear function  $\tau : \mathcal{W} \mapsto \mathbb{C}$ , the free cumulants are multilinear maps  $\kappa_n : \mathcal{W}^n \mapsto \mathbb{C}$  defined recursively via

$$\tau(a_1 \cdots a_n) = \sum_{\pi \in \text{NC}_n} \prod_{V \in \pi} \kappa_{|V|}(\{a_i\}_{i \in V}) \quad (2.20)$$

where  $V$  is a block of a NC partition  $\pi$  and  $|V|$  denotes its cardinality.

We shall often drop free and just call them cumulants. We refer (2.20) as the *moment-cumulant formula*. There is only one term in the sum involving the highest cumulant  $\kappa_n$ , thus the formula can inductively be inverted for the cumulants in terms of the moments.<sup>30</sup> Therefore, (2.20) is indeed the defining formula for free cumulants.

We are often interested in the relation between the moments and cumulants of a single random variable  $a$ , we can define the abbreviations for the  $n$ -th moment and  $n$ -th cumulant of  $a$  as

$$m_n(a) := \tau(a^n) \quad \kappa_n(a) := \kappa_n(a, \dots, a) . \quad (2.21)$$

We then have a special case of the free moment-cumulant formula (2.20),

$$m_n(a) = \sum_{\pi \in \text{NC}_n} \prod_{V \in \pi} \kappa_{|V|}(a) . \quad (2.22)$$

We shall remark that for a single self-adjoint operator  $a$ , the moments  $m_n(a)$  are identical to the classical moments of a random variable  $X$  with the probability distribution given by the spectral distribution of  $a$ . Because of the distinction between freeness and independence, however, the corresponding classical cumulants, denoted as  $u_n(X)$ , are defined differently from  $\kappa_n(a)$ . The same formula (2.22) is used up to changing the NC partitions to all partitions of  $[n]$ , denoted as  $\mathcal{P}_n$ , yielding the classical moment-cumulant formula,

$$m_n(X) = \sum_{\pi \in \mathcal{P}_n} \prod_{V \in \pi} u_{|V|}(X) . \quad (2.23)$$

Hence, one can see that  $\kappa_n$  and  $u_n$  with  $n \geq 4$  are different from each other. The classical cumulants  $u_n(X)$  are also known as the Ursell functions in statistic physics or the connected correlation functions in QFT. In a Feynman diagram expansion of a partition function in QFT, the  $n^{\text{th}}$  cumulant

<sup>27</sup>There is a natural partial order defined for NC partitions, called the reverse refinement order. We say  $\pi_1 \leq \pi_2$  if each block of  $\pi_1$  is contained in the blocks of  $\pi_2$ .

<sup>28</sup>The corresponding permutation of  $\bar{\pi}$  is equivalently given by  $\pi^{-1} \eta$  where  $\eta$  is the cyclic permutation [53].

<sup>29</sup>One can define a metric on the permutations known as the *Cayley distance*,  $d(\pi_1, \pi_2) := n - |\pi_1^{-1} \pi_2|$ , which also counts the minimal number of transpositions needed to transform  $\pi_1$  to  $\pi_2$ . Then equation  $|\pi_1| + |\pi_2| \leq |\pi_1 \pi_2| + n$  follows from the triangle inequality of the Cayley distance.

<sup>30</sup>More precisely, the set of non-crossing partitions forms a lattice with respect to refinement order and the cumulants are given by the Möbius inversion with respect to this order.

$u_n(\phi)$  defined for a field  $\phi$  gives the  $n$ -connected Green functions that connect exactly  $n$  sources. A cumulant  $u_n(\phi)$  should thus be understood as capturing genuine  $n$ -partite interactions, as opposed to independence or correlations that can be decomposed into interactions among fewer parties. Analogously, we will see that the free cumulants are precisely computed by the replica wormholes in the planar gravitational saddles of the replica trick GPI.

Here comes the punchline: *Freeness is equivalent to vanishing mixed cumulants.*

**Theorem 1** (Speicher [54]). *The fact that  $a$  and  $b$  in  $(\mathcal{W}, \tau)$  are free is equivalent to vanishing  $\kappa_n(x_1, \dots, x_n)$  whenever  $n \geq 2$ ,  $x_i = a$  or  $b \ \forall i$  and  $x_i = a, x_j = b$  for some  $i, j$ .*

It's now clear that any joint moments of free random variables can be written in terms of the individual cumulants, because the mixed cumulants vanish. Since each individual cumulant is a polynomial of the individual moments via inverting the moment-cumulant formula, we have established the desired property that the joint moments are polynomials of the individual moments for free random variables. This drastic simplification is what freeness buys us.

As discussed previously, from the individual moments of a collection of free self-adjoint (or normal cf. Footnote 17) random variables, we can extract the spectral distribution of each random variable. Their freeness further allows us to extract the analytic distributions for some combinations of free random variables via tools that we will now come to.

## 2.4 Free harmonic analysis

We now introduce some tools from free harmonic analysis that will prove useful.

### 2.4.1 The Cauchy transform

We start with the *Cauchy-Stieltjes transform*,<sup>31</sup> or simply the Cauchy transform, of a distribution  $\mu$  on  $\mathbb{R}$ ,

$$G_\mu(z) := \int_I \frac{d\mu(x)}{z - \lambda}, \quad z \in \mathbb{C} \setminus I, \quad (2.24)$$

where  $I$  denotes the support of  $\mu$  on the real line.

The Cauchy transform is defined on both the upper and lower complex half-plane, denoted as  $\mathbb{C}^+ := \{z : \text{Im}(z) > 0\}$  and  $\mathbb{C}^- := \{z : \text{Im}(z) < 0\}$  respectively. However, note that  $G_\mu(z)^* = G_\mu(z^*)$  so we only need to focus on the domain  $\mathbb{C}^+$ . In particular, it is a holomorphic function from  $\mathbb{C}^+$  to  $\mathbb{C}^-$ . We can extract the spectral distribution  $\mu(x)$  via the inverse transform, known as the *Stieltjes inversion*,

$$\mu(x) = -\frac{1}{\pi} \lim_{\epsilon \rightarrow 0} \text{Im} G_\mu(x + i\epsilon). \quad (2.25)$$

Consider now a self-adjoint operator  $a \in (\mathcal{W}, \tau)$ . The Cauchy transform of its spectral distribution can be regarded as a moment generating function for  $a$ ,

$$G_{\nu_a}(z) = \sum_{n=0}^{\infty} \frac{m_n(a)}{z^{n+1}}. \quad (2.26)$$

In the random matrix theory, the Cauchy transform for the spectral distribution of a random matrix  $A \in (L^\infty(\Omega) \otimes M_N(\mathbb{C}), \mathbb{E} \otimes \text{tr})$  is also known as the (trace of) resolvent of  $A$ .

<sup>31</sup>The Stieltjes transform usually refers to the Cauchy transform with a minus sign.

### 2.4.2 Free additive convolution and $R$ -transform.

Freeness allows us to break down the moments of any polynomial of several free random variables in terms of the cumulants/moments of each individual random variable. It's natural to ask about the spectral distribution  $\nu_{a+b}$  of the sum of two free self-adjoint random variables  $a$  and  $b$ . This distribution is known as the *free additive convolution* of  $\nu_a$  and  $\nu_b$  [46],

$$\nu_{a+b} = \nu_a \boxplus \nu_b \quad (2.27)$$

where the notation is chosen to indicate that this spectral distribution can be obtained from their spectral distributions  $\nu_a$  and  $\nu_b$  thanks to the freeness. The free additive convolution  $\boxplus$  is associative and commutative operation. Note this operation is defined here for any two compactly supported distributions on  $\mathbb{R}$ ,<sup>32</sup> Therefore, it only depends on the spectral distributions  $\nu_a$  and  $\nu_b$  rather than the specific variables  $a$  and  $b$ , because different random variables can share the same spectrum.

In order to extract the convoluted distribution, we look at its cumulants,  $\kappa_n(a+b)$ . Recall that free cumulants are multilinear maps and the mixed free cumulants for free random variables vanish, from which it immediately follows that, for any  $n$ ,

$$\kappa_n(a+b) = \kappa_n(a) + \kappa_n(b) . \quad (2.28)$$

Therefore, we obtain a simple relation between  $a+b$  and the individual  $a$  and  $b$  in terms of the free cumulants. Now we need to make an analytic connection back to the moments via the  $R$ -transform. For a deterministic spectral distribution  $\nu_a$ , its  $R$ -transform is defined as

$$R_{\nu_a}(z) = \sum_{n=1}^{\infty} \kappa_n(a) z^{n-1} , \quad (2.29)$$

The  $R$ -transform can be viewed as a free cumulant generating function. Note that unlike the Cauchy transform that is analytic on the entire  $\mathbb{C}^+$ , the  $R$ -transform is usually only analytical for small  $z$  and the domain depends on the measure  $\nu_a$ . See Chapter 3 in Mingo-Speicher [21] for a detailed account of its analytic domain.

Then (2.28) implies the convolution  $\boxplus$  becomes additive under the  $R$ -transform,

$$R_{\nu_a \boxplus \nu_b}(z) = R_{\nu_a}(z) + R_{\nu_b}(z) . \quad (2.30)$$

Recall that the Cauchy transform (2.24) is like a moment generating function. Speicher showed that it is related to  $R$ -transform via [54],

$$R_{\nu_a}(G_{\nu_a}(z)) + G_{\nu_a}(z)^{-1} = z . \quad (2.31)$$

Treating  $G_{\nu_a}$  as a variable  $z$ , we can write it as

$$R_{\nu_a}(z) + z^{-1} = G_{\nu_a}^{-1}(z) \quad (2.32)$$

where  $G_{\nu_a}^{-1}(z)$  is the inverse function of the Cauchy transform. It can be further re-written as

$$G_{\nu_a}(R_{\nu_a}(z) + 1/z) = z . \quad (2.33)$$

Hence, the Cauchy transform is the functional inverse of  $R_{\nu_a}(z) + 1/z$ .<sup>33</sup>

The equation (2.31) connects the  $R$ -transform and the Cauchy transform and thus the spectral distribution. Let's summarize the algorithm for free additive convolution:

<sup>32</sup>It can also be defined for distributions with non-compact support on  $\mathbb{R}$  [55].

<sup>33</sup>This is sometimes used as an alternative definition for the  $R$ -transform.

**Algorithm 1.** *Free additive convolution  $\nu_a \boxplus \nu_b$  via the  $R$ -transform:*

1. compute the Cauchy transform from the spectrum measures  $\nu_a$  and  $\nu_b$  using (2.24);
2. compute the  $R$ -transforms using (2.31) and apply (2.30) to obtain  $R_{\nu_a \boxplus \nu_b}$ ;
3. find the Cauchy transform and then extract  $\nu_a \boxplus \nu_b$  via the Stieltjes inversion (2.25).

### 2.4.3 Free multiplicative convolution and $S$ -transform.

Similarly, one can also find the spectral distribution of the product of two free positive self-adjoint random variables  $a$  and  $b$ ,  $ab$  or  $ba$ . In particular, let's consider the positive operators  $\sqrt{b}a\sqrt{b}$  or  $\sqrt{a}b\sqrt{a}$  that share the same spectrum. It is given by the *free multiplicative convolution* [49],

$$\nu_{\sqrt{ab}\sqrt{a}} = \nu_{\sqrt{ba}\sqrt{b}} = \nu_a \boxtimes \nu_b \quad (2.34)$$

where the notation is chosen to indicate that this spectral distribution can be obtained from their spectral distributions  $\nu_a$  and  $\nu_b$  thanks to the freeness. Just like free additive convolution, the free multiplicative convolution  $\boxtimes$  is an associative and commutative operation. Note this operation is defined for any two compactly supported distributions on  $\mathbb{R}_+$ .<sup>34</sup> Therefore, it only depends on the spectral distributions  $\nu_a$  and  $\nu_b$  rather than the specific variables  $a$  and  $b$ . The convoluted distribution has bounded support if  $\nu_a$  and  $\nu_b$  have bounded support.

Let's see how this works with the help of the  $S$ -transform. Analogous to the  $R$ -transform, which can help us calculate the free additive convolution, the free convolution can be computed by the  $S$ -transform, where the convolution turns into a multiplication. We first define another moment function  $\omega$  just as an intermediate tool,

$$\omega_\mu(z) := zG_\mu(z) - 1 = \sum_{n=1}^{\infty} \frac{m_n(\mu)}{z^n}. \quad (2.35)$$

The  $S$ -transform  $S_\mu(z)$  is then defined as

$$S_\mu(z) := \frac{1+z}{z\omega_\mu^{-1}(z)} \quad (2.36)$$

where  $\omega_\mu^{-1}(z)$  is the inverse function of (2.35). Generally, the  $S$ -transform is only analytic in the neighbourhood of 0. See for instance [55] for more details its analytic domain. We shall also note that the  $S$ -transform and  $R$ -transform are related via the composition inverse of the functions

$$x = zR_\mu(z), \quad z = xS_\mu(x). \quad (2.37)$$

Using this, one can transfer between the  $R$ -transform and the  $S$ -transform, and it shall prove very useful later.

The  $S$ -transform allows us to compute the free multiplicative convolution with

$$S_{\nu_a \boxtimes \nu_b}(\omega) = S_{\nu_a}(\omega)S_{\nu_b}(\omega). \quad (2.38)$$

We then invert (2.36) to obtain  $\omega_{\nu_a \boxtimes \nu_b}(z)$ , thus  $G_{\nu_a \boxtimes \nu_b}(z) = (\omega_{\nu_a \boxtimes \nu_b}(z) + 1)/z$  and finally the spectral distribution  $\nu_a \boxtimes \nu_b$  via the Stieltjes inversion (2.25). Let's summarize the algorithm for the free multiplicative convolution,

---

<sup>34</sup>It can also be defined for distributions with non-compact support on  $\mathbb{R}_+$  [55] or distributions supported on the unit circle in  $\mathbb{C}^+$  [50]. One can also relax the condition that one distribution can be supported on  $\mathbb{R}$  provided the other distribution is supported  $\mathbb{R}_+$  [56].

**Algorithm 2.** Free multiplicative convolution  $\nu_a \boxtimes \nu_b$  via the  $S$ -transform:

1. compute the moment functions (2.35)  $\omega_{\nu_a}$  and  $\omega_{\nu_b}$  from the Cauchy transforms (2.24);
2. compute the  $S$ -transforms using (2.36) and apply (2.38) to obtain  $S_{\nu_a \boxtimes \nu_b}$ ;
3. find the Cauchy transform and then extract  $\nu_a \boxtimes \nu_b$  via the Stieltjes inversion (2.25).

Above, we took Voiculescu’s approach to introduce the free multiplicative convolution in terms of the  $S$ -transforms [49]. The following result due to Nica-Speicher provides a complementary combinatorial perspective on the free multiplicative convolution [20]. Let  $(\mathcal{W}, \tau)$  be a non-commutative probability space and  $a, b \in (\mathcal{W}, \tau)$  such that they are free. We have the moments of the product  $ab$ ,

$$m_n(ab) = \sum_{\pi \in \text{NC}_n} \left( \prod_{V \in \pi} \kappa_{|V|}(a) \cdot \prod_{\bar{V} \in \bar{\pi}} m_{|\bar{V}|}(b) \right), \quad (2.39)$$

and free cumulants

$$\kappa_n(ab) = \sum_{\pi \in \text{NC}_n} \left( \prod_{V \in \pi} \kappa_{|V|}(a) \cdot \prod_{\bar{V} \in \bar{\pi}} \kappa_{|\bar{V}|}(b) \right). \quad (2.40)$$

Notice that these formulas look exactly like what the replica trick GPI gives us (1.8). This is the key observation of this work and we shall match the details later in Section 3.1.

In practice, the free multiplicative convolution and the free additive convolution via the  $S, R$ -transforms can be hard to evaluate either analytically or even numerically. The difficulty lies in dealing with the functional inverses when converting the  $S, R$ -transforms back to the Cauchy transforms.<sup>35</sup> In fact, one can circumvent from dealing with the  $S, R$ -transforms directly and instead resort to the subordinate functions [59]. They have better analytic properties and boil down to fixed-point equations that are numerically tractable. (cf. Chapter 5 in [22] for an introduction).

There is a general procedure developed to evaluate any polynomials of a collection of free random variables. (cf. Chapter 10 in Mingo-Speicher [21] and [60] for details.) The procedure builds on the operator-valued free probability theory [61], that generalizes the scalar-valued case that we’ve been using. It then uses a linearization trick to formulate any polynomial as a linear combination of free variables with matrix coefficients, turning the problem into an addition of operator-valued free random variables [62, 63]. Then, instead of the  $R$ -transforms, one can implement the subordination formalism to compute the sum [64]. Later it turns out that we do not have to implement this general machinery for our problem.

#### 2.4.4 Free compression

Consider a projection  $p \in (\mathcal{W}, \tau)$  of trace  $\tau(p) < 1$  acting as  $pap$  on  $a \in (\mathcal{W}, \tau)$ . It has the spectral distribution  $\tau(p)\delta_1 + (1 - \tau(p))\delta_0$ . These projected random variables form a subalgebra  $(p\mathcal{W}p, \tau_p) \subset (\mathcal{W}, \tau)$  where  $\tau_p := \tau/\tau(p)$  is the renormalized trace and  $p$  is the unit element of  $p\mathcal{W}p$ . This projected subalgebra forms a non-commutative probability space and is referred as a *compression* of  $(\mathcal{W}, \tau)$ . The compression inherits the structural properties of  $(\mathcal{W}, \tau)$ . For instance, if  $(\mathcal{W}, \tau)$  is a  $W^*$ -probability space, so is  $(p\mathcal{W}p, \tau_p)$ . In particular, if  $(\mathcal{W}, \tau)$  is a Type II<sub>1</sub> factor, so is  $(p\mathcal{W}p, \tau_p)$ .

Consider now that  $p$  is free from some  $a \in (\mathcal{W}, \tau)$ . We call  $pap$  a *free compression* of  $a$  by  $p$ . We are interested in applying the free multiplicative convolution to compute the spectral distribution,

<sup>35</sup>There are some efforts from the machine learning community in developing algorithms to directly handle inverting the  $S$ -transforms [57, 58].



$\nu_{pap/\tau(p)}$ , of  $pap/\tau(p) \in (\mathcal{W}, \tau)$ , where it's customary to normalize the compressed variable by  $\tau(p)$ . One can show that

$$\nu_{pap/\tau(p)}^p = \nu_a \boxtimes (\tau(p)\delta_{1/\tau(p)} + (1 - \tau(p))\delta_0) = \tau(p)\nu_a^{\boxplus 1/\tau(p)} + (1 - \tau(p))\delta_0 \quad (2.41)$$

where the  $\boxplus$  power denotes a  $1/\tau(p)$ -fold free additive convolution, provided  $1/\tau(p)$  is an integer. If we restrict to the subalgebra  $(p\mathcal{W}p, \tau_p)$  that the random variable is compressed to, we switch to the renormalized trace  $\tau_p$  and the corresponding spectral distribution, denoted as  $\nu_{pap/\tau(p)}^p$ , is simply

$$\nu_{pap/\tau(p)}^p = \nu_a^{\boxplus 1/\tau(p)}. \quad (2.42)$$

Surprisingly, we see that a free compression  $p$  essentially implements a  $1/\tau(p)$ -fold free additive convolution up to a rescaling or  $\tau(p)$ . In particular, we have the following relation for the cumulants defined w.r.t.  $\tau_p$ ,

$$\tau(p)\kappa_n(pap/\tau(p)) = \kappa_n(a), \quad \forall n \in \mathbb{N}, \quad (2.43)$$

and thus the  $R$ -transforms satisfy

$$\tau(p)R_{\nu_{pap/\tau(p)}^p} = R_{\nu_a}. \quad (2.44)$$

It's also useful to write (2.42) as

$$\nu_{pap}^p = \nu_a^{\boxplus 1/\tau(p)} \circ D_{\tau(p)}. \quad (2.45)$$

where  $D_{\tau(p)}$  denotes a dilation (rescaling) of the distribution. This operation is defined to rescale any distribution  $\nu(x)$  as

$$\nu \circ D_\lambda(x) := \lambda\nu(\lambda x). \quad (2.46)$$

Equation (2.45) follows from the general fact that if one rescales a self-adjoint operator by  $a$ , then its spectral distribution is dilated by  $1/\lambda$ ,

$$\nu_{\lambda a} = \nu_a \circ D_{1/\lambda}. \quad (2.47)$$

In fact, we can use (2.42) to extend the definition of the free additive convolution power to any real number  $t \geq 1$  with  $\tau(p) = 1/t$ .<sup>36</sup> Moreover, one can use the free compression to generate, from a given compactly supported  $\nu$ , a  $\boxplus$ -semigroup<sup>37</sup> of compactly supported probability measures [66]. It is denoted as,  $(\nu_t)_{t \geq 1}$ , and it satisfies

$$\nu_{t+s} = \nu_t \boxplus \nu_s, \quad \nu_1 = \nu. \quad (2.48)$$

Here is how the  $\boxplus$ -semigroup can be constructed. Starting from a given  $\nu$ , we assign a self-adjoint random variable  $a$  with this spectral distribution. We also introduce a projection of trace  $1/t$  and put  $a$  and  $p$  in a common probability space  $(\mathcal{W}, \tau)$  under the free product such that  $p$  and  $a$  are free. Then we consider  $pap \in (p\mathcal{W}p, \tau_p)$  and define  $\nu_t := \nu_{tpap}^p = \nu^{\boxplus t}$ . The semigroup property (2.48) is manifestly satisfied. We will later see an important family of distributions for which the semigroup extends also to  $0 < t < 1$ .

<sup>36</sup>This is possible, for instance, in a Type  $\text{II}_1$  factor which has no minimal projection.

<sup>37</sup>See [65] for a free multiplicative counterpart.

## 2.5 Circular, semi-circular, quarter-circular and free Poisson distributions

In this subsection, we introduce some canonical examples of probability distributions in free probability theory that are relevant for our discussions later. These are closely related to some canonical random matrix ensembles and we shall mention some in passing. The details on random matrices are provided in Appendix B.

One of the most important and random matrix ensemble is the Gaussian Unitary Ensemble (GUE) of Hermitian matrices. An  $N \times N$  GUE matrix  $S_N$  have i.i.d. complex Gaussian matrix elements with the real and complex parts independently distributed with zero mean and variance  $1/2N$ , and the diagonal are i.i.d. Gaussian-distributed with zero mean and variance  $1/N$ . The spectral distributions of a sequence of GUE  $(S_N)_N$  converges almost surely to the semi-circular distribution,

$$\mu_s(x) := (2\pi)^{-1} \sqrt{(4-x^2)} \quad (2.49)$$

with support  $x \in [-2, +2]$ . We call any element  $s \in (\mathcal{W}, \tau)$  a *semi-circular element* if its spectral distribution  $\mu_s$  is given by the semi-circular distribution. Its moments are given by the Catalan numbers,

$$m_n(s) = \lim_{N \rightarrow \infty} \mathbb{E} \operatorname{tr}(S_N^n) = \frac{1}{n+1} \binom{2n}{n} \text{ for even } n, \text{ and zero otherwise.} \quad (2.50)$$

Its free cumulants are given by

$$\kappa_n(s) = \delta_{n,2}. \quad (2.51)$$

The semi-circular distribution is thus most easily characterized by that its only non-zero cumulant is  $\kappa_2$ . This fact directly leads to the *free central limit theorem*. The classical theorem says that for a collection of  $n$  independent random variables  $x_1, \dots, x_n$  with zero mean and unit variance. Their average  $(\sum_{i=1}^n x_i)/\sqrt{n}$  converges (in distribution) to a standard normal random variable. The free probability counterpart is the free central limit theorem, which asserts that for a collection of free self-adjoint random variables  $\{a_i\}_{i=1}^n$  in  $(\mathcal{W}, \tau)$  with zero mean and unit variance ( $\kappa_2(a_i) = 1$ ), their average  $(\sum_{i=1}^n a_i)/\sqrt{n}$  converges (in distribution) to a *semicircular element*  $s \in (\mathcal{W}, \tau)$ .<sup>38</sup>

A *circular element*  $c \in (\mathcal{W}, \tau)$  is defined to be

$$c := \frac{s_1 + is_2}{\sqrt{2}} \quad (2.52)$$

for two free semi-circulars  $s_1, s_2$ . Its spectral distribution is the uniform distribution over the unit disk on the complex plane, called the *circular distribution*,

$$\mu_c(z) := \pi^{-1} \mathbf{1}_{|z| \leq 1}. \quad (2.53)$$

They can be modelled by large random matrix ensemble  $C_N$  of form

$$C_N = S_N^{(1)} + iS_N^{(2)} \quad (2.54)$$

---

<sup>38</sup>We can sketch here the proof for this theorem. The spectral distribution of the addition of a collection of free random variables is described by the free additive convolution, under which the free cumulants add. Therefore, all the free cumulants have linear growth in  $n$  whereas the denominator  $\sqrt{n}$  suppresses  $\sum_i^n \kappa_m(a_i)$  by  $n^{m/2}$ , so only the first ( $m = 1$ ) and second cumulant ( $m = 2$ ) survive in the limit  $n \rightarrow \infty$ . Since the first cumulant equals to the first moment which vanishes for all  $a_i$ , we have only a non-zero  $\kappa_2 = 1$ . Hence, the limiting distribution is semicircular and the free central limit theorem follows.

for two  $N \times N$  independent GUE matrices  $S_N^{(1)}$  and  $S_N^{(2)}$ .  $C_N$  is called the (square) *Ginibre ensemble* and they have i.i.d. matrix elements that follow the complex Gaussian distribution of zero mean and variance  $1/N$ . Its limiting spectral distribution is given by the circular distribution. The Ginibre ensemble has an important property that its joint distribution of the matrix elements are invariant under any unitary action from either left or right. We say it is both *left and right unitarily invariant*.<sup>39</sup>

Now consider the modulus of a circular element, known as a *quarter-circular element*

$$q := |c| := \sqrt{cc^*} . \quad (2.55)$$

Its spectral distribution follows the *quarter-circular distribution*,

$$\mu_q(x) := \pi^{-1} \sqrt{(4-x^2)} \quad (2.56)$$

with support  $x \in [0, 2]$ . The spectral distribution is also shared by the modulus of a semi-circular element  $|s|$ . Its moments are given by  $m_n(q) = \frac{2^n \Gamma(1/2+n/2)}{\sqrt{\pi} \Gamma(2+n/2)}$ . It's more interesting to look at the moments of its square, which are given by the Catalan numbers,

$$m_n(q^2) = \frac{1}{n+1} \binom{2n}{n} , \quad \forall n \quad (2.57)$$

and the cumulants are given by

$$\kappa_n(q^2) = 1 , \quad \forall n . \quad (2.58)$$

A useful lemma due to Voiculescu claims that a circular element admits the polar decomposition [28],

$$c = uq \quad (2.59)$$

in terms of a Haar unitary  $u$  and a quarter-circular element  $q$  that are free from each other. An element  $u \in (\mathcal{W}, \tau)$  is called a Haar unitary if  $u^*u = uu^* = 1$  and  $\tau(u^n) = \delta_{0,n}, \forall n \in \mathbb{N}$ . Its random matrix counterpart is of course the Haar unitary ensemble, also known as the circular unitary ensemble.

The quarter-circular distribution (2.56) is a special instance of the *free Poisson distribution*. Let us motivate it with the *free Poisson limit theorem*. It is the non-commutative analog to the Poisson limit theorem that asserts the sum of Bernoulli random variables (coin flips) asymptotes to a Poisson random variable. The free Poisson distribution can be obtained from multiple free additive convolutions of the Bernoulli distribution, which has two outcomes 0 and  $\alpha$ . Then  $\alpha$  is the jump size and  $\lambda$  is the rate of obtaining the outcome  $\alpha$ . We have<sup>40</sup>

$$\mu_{\lambda, \alpha} = \lim_{N \rightarrow \infty} [(1 - \lambda/N)\delta_0 + (\lambda/N)\delta_\alpha]^{\boxplus N} , \quad (2.60)$$

which is known as the *free Poisson distribution* and it admits a close-form expression,

$$\mu_{\lambda, \alpha} := \begin{cases} (1 - \lambda)\delta_0 + \tilde{\mu}, & \text{if } 0 \leq \lambda \leq 1 \\ \tilde{\mu}, & \text{if } \lambda > 1 \end{cases} \quad (2.61)$$

where  $\tilde{\mu}(x) := \frac{1}{2\pi\alpha x} \sqrt{(x_+ - x)(x - x_-)} \mathbf{1}_{[x_-, x_+]}$ ,  $x_\pm := \alpha(1 \pm \sqrt{\lambda})^2$ .

<sup>39</sup>See, for instance, Lemma 1 in [67].

<sup>40</sup>The classical Poisson distribution is obtained by taking the classical  $*$ -convolution power.

Its moments are given by

$$m_n(\mu_{\lambda,\alpha}) = \alpha^n \sum_{k=1}^n \lambda^{n-k} N(n, k) \quad (2.62)$$

where  $N(n, k) := \binom{n}{k} \binom{n}{k-1} / n$  are the Narayana numbers. When  $\lambda = \alpha = 1$ , the moments reduce to the Catalan numbers of the quarter-circular distribution. The free Poisson distribution is most concisely characterized by its cumulants,

$$\kappa_n(\mu_{\lambda,\alpha}) = \lambda \alpha^n . \quad (2.63)$$

The free Poisson distribution is also known as the Marchenko-Pastur (MP) distribution under a change of variables. The MP distribution describes the limiting spectrum of a Wishart ensemble which has the form  $X_N X_N^*$  where  $X_N$  is a (rectangular)  $N \times M$  Ginibre ensemble with i.i.d. matrix entries of zero mean and variance  $\alpha/M$  and  $M/N \rightarrow \lambda$ .

Consider a generalization of the free Poisson limit theorem such that the jump size is not a deterministic value  $\alpha$ , but rather follows some probability distribution  $\nu$ ,

$$\mu_{\lambda,\nu} = \lim_{N \rightarrow \infty} [(1 - \lambda/N)\delta_0 + (\lambda/N)\nu]^{\boxplus N} . \quad (2.64)$$

This is known as the *free compound Poisson distribution*. They are most easily characterized by the cumulants.

$$\kappa_n(\mu_{\lambda,\nu}) = \lambda m_n(\nu) . \quad (2.65)$$

A special case of a compound Poisson distribution is when  $\lambda = 1$ . Then the cumulants of  $\mu_{\lambda,\nu}$  are exactly given by the moments of  $\nu$ . In this case, suppose  $\nu$  is the spectral distribution of some random variable  $a$ , then  $\mu_{\lambda,\nu}$  is the spectral distribution of  $cac^*$ . Namely, the conjugation by circular elements swaps moments to cumulants. To see how, consider the general moment formula of the free multiplicative convolution between  $a$  and  $cc^*$ , which is the square of a quarter-circular element  $q^2$ . Consider the general moment formula of the free multiplicative convolution (2.39),

$$m_n(cac^*) = \sum_{\pi \in \text{NC}_n} \left( \prod_{V \in \pi} \kappa_{|V|}(q^2) \cdot \prod_{\tilde{V} \in \tilde{\pi}} m_{|\tilde{V}|}(a) \right) = \sum_{\pi \in \text{NC}_n} \prod_{\tilde{V} \in \tilde{\pi}} m_{|\tilde{V}|}(a) = \sum_{\pi \in \text{NC}_n} \prod_{V \in \pi} m_{|V|}(a) , \quad (2.66)$$

which implies (2.65) via the definition of free cumulants (2.22).

The natural random matrix sequence with the limiting spectral distribution of a free compound Poisson distribution is the sample covariance matrix ensemble,<sup>41</sup>

$$\sqrt{A_N} X_N X_N^* \sqrt{A_N} , \quad (2.67)$$

where  $X_N$  is an  $N \times M$  Ginibre ensemble with i.i.d. matrix entries of zero mean and variance  $1/M$ .  $A_N$  is a sequence of independent positive semidefinite matrices  $A_N$  with the limiting spectral distribution  $\nu$ .

---

<sup>41</sup> In statistics,  $A_N$  represents the covariance matrix of a centered random vector  $\mathbf{y}$  with  $N$  data entries,  $A_N = \mathbb{E} \mathbf{y} \mathbf{y}^*$ , and we can write the random vector as  $\mathbf{y} = \sqrt{A_N} \mathbf{x}$  where  $\mathbf{x}$  is some vector with centered i.i.d. entries with unit variance. A sample covariance matrix  $M^{-1} \sum_i^M \mathbf{y}_i \mathbf{y}_i^*$  is an estimator for  $A_N$  with  $M$  i.i.d. draws of  $\mathbf{y}$ . Alternatively, we can write the sample covariance matrix as (2.67).

A compound Poisson distribution is a canonical example of *free infinitely divisible distributions*, which are probability distributions that can be decomposed as<sup>42</sup>

$$\mu = \mu_{1/k}^{\boxplus k} \quad (2.68)$$

for any  $k \in \mathbb{N}$ . We can then make use of  $\mu_{1/k}$  to define any fractional convolution power  $\mu^{\boxplus l/k}$  as  $\mu_{1/k}^{\boxplus l}$ . By continuity, we obtain a one-parameter family of spectral distributions,  $(\mu_t)_{t>0}$ , that actually forms a  $\boxplus$ -semigroup under the associative action of the free multiplicative convolution  $\boxplus$ ,  $\mu_{s+t} = \mu_s \boxplus \mu_t$ .

Conversely, given a such a semigroup  $(\mu_t)_{t>0}$ , any element  $\mu_t$  is infinitely divisible. Hence, free infinite divisibility is equivalent to the existence of a  $\boxplus$ -semigroup for  $t \geq 0$ . Therefore, we see that while every compactly supported probability measure belongs to a  $\boxplus$ -semigroup with  $t \geq 1$ , the ones with the extended parameter range  $t > 0$  are very special. See Chapter 3 in Hiai-Petz [68] for a good discussion of free infinite divisibility.

It is then evident that the  $R$ -transforms of  $\mu_t$  are simply multiples of the  $R$ -transform of  $\mu_1 = \mu$ ,

$$R_{\mu_t} = tR_\mu, \quad t \geq 0. \quad (2.69)$$

It turns out that one can be more explicit about the  $R$ -transform of an infinitely divisible  $\mu$ . It admits the following form, known as the free Lévy-Khintchine formula,<sup>43</sup>

$$R_\mu(z) = \kappa_1(\mu) + \int_{\mathbb{R}} \frac{z}{1-xz} \eta(x) dx, \quad \forall z \in \mathbb{C} \setminus \mathbb{R} \cup \{0\}, \quad (2.70)$$

for some positive finite measure  $\eta$  on  $\mathbb{R}$ . Rather than being analytic only for small  $|z|$ , such  $R_\mu(z)$  is analytic on both  $\mathbb{C}^\pm$ , and it maps  $\mathbb{C}^\pm \rightarrow \mathbb{C}^\pm$ .

In particular, any compound Poisson distribution  $\mu_{\lambda,\nu}$  can be decomposed as

$$\mu_{\lambda,\nu} = \mu_{\lambda/k,\nu}^{\boxplus k} \quad (2.71)$$

and the semigroup of distributions is defined as  $\mu_t := \mu_{t\lambda,\nu}$ . Using the cumulants (2.65). Its  $R$ -transform reads

$$R_{\mu_{\lambda,\nu}}(z) = \lambda x \sum_{n=0}^{\infty} \int x^n z^n \nu(x) dx = \int \frac{\lambda x}{1-xz} \nu(x) dx. \quad (2.72)$$

which admits the form of the Lévy-Khintchine formula (2.70) with  $\kappa_1 = \lambda m_1(\nu)$  and  $\eta(x) = x^2 \lambda \nu(x)$ .

In sum, thanks to the free infinite divisibility, the  $R$ -transforms of the free compound Poisson distributions are relatively simple to work with. Unlike the free Poisson distribution, a free compound Poisson distribution doesn't generally have a closed-form expression, but a closed-form expression for the  $R$ -transform is already good enough for many purposes.

### 3 Results

We now come to the main observation that the replica trick GPI should be understood as a free multiplicative convolution of two probability distributions, one of which is determined by the reservoir Hamiltonian or any quantum processing we apply to the Hawking radiation collected, and the other is

<sup>42</sup>A similar notion of infinite divisibility also exists for the free multiplicative convolution.

<sup>43</sup>This is the free analog of the classical Lévy-Khintchine formula that characterizes the characteristic functions of infinitely divisible distributions.

determined by the JT gravity+EOW brane theory. We need to sort out some technical issues in order to make them match. Then we use the tools from free harmonic analysis to evaluate the convolution.

We shall also infer from the convolution formula a random matrix model that allows us to compute any spectral functions of the radiation density matrix. Surprisingly, it fits perfectly with Page's original proposal addressing how an evaporating black hole should be treated information-theoretically. To support the validity of the free convolution formula, we perform some consistency checks with some known results in the literature. Finally, we show how to re-formulate the entropy formula such that it is free of free probability, but resembles Wall's formula for the generalized entropy.

### 3.1 Replica trick as a free multiplicative convolution

We would like to match (1.8) with (2.39). The main idea is that the collection of gravitational partition functions  $\{Z_n\}$ , obtained from the connected replica wormholes, should be viewed as free cumulants rather than moments.

These  $Z_n$  partition functions are defined by the boundary condition that consists of alternative JT asymptotic boundaries and the EOW branes. They are computed by PSSY based on an earlier calculation by Yang [69]. The results read

$$Z_n = \int_0^\infty dE \rho(E) y(E)^n, \quad y(E) := e^{-\beta E} 2^{1-2\mu} |\Gamma(\mu - 1/2 + i\sqrt{2E})|^2, \rho(E) := e^{S_0} \sinh(2\pi\sqrt{2E}) / (2\pi^2) \quad (3.1)$$

where  $y(E)$  is a Boltzmann-type factor that is influenced by the EOW brane, and  $\rho(E)$  is the density of states. After normalization with  $Z_1^n$ . We can rewrite it as the  $n^{\text{th}}$  moment of some distribution,

$$\frac{Z_n}{Z_1^n} = \int_0^{y(0)} \frac{\rho(\tilde{E}(y))}{-y'(\tilde{E}(y))} (y/Z_1)^n dy = \int_0^{x(0)} \frac{\rho(E(x))}{-y'(E(x))} x^n dx =: \int_0^{x(0)} \nu_b(x) x^n dx \quad (3.2)$$

where  $\tilde{E}(y)$  is the inverse function of  $y(E)$ ,  $x := y/Z_1$ ,  $E(x) = \tilde{E}(xZ_1)$  in the second equality and lastly we've defined a distribution function

$$\nu_b(x) := \frac{\rho(E(x))}{-y'(E(x))}. \quad (3.3)$$

Note that  $y' < 0$  as  $y(E)$  is a decreasing function in  $E$ . Note that  $\rho(E)$  is not integrable so is  $\nu_b$  whose zeroth moment diverges at  $x \rightarrow 0$ , so  $\nu_b$  is *not* a probability distribution, but only a positive measure over  $[0, x(0)]$ .

The Bekenstein-Hawking entropy of the JT black hole + EOW brane system is given by

$$S_{\text{BH}} := -\partial_N (Z_n/Z_1^n)|_{n=1} = -\int_0^{x(0)} x \log x \nu_b(x) dx. \quad (3.4)$$

In other words, the BH entropy is the entropy of the measure  $\nu_b$ . In particular,  $\mu \gg 1/\beta$ , we have

$$S_{\text{BH}} = S_0 + 4\pi^2/\beta + \mathcal{O}(1). \quad (3.5)$$

It's more handy to work with a probability distribution, so we let  $\nu_{b_N}$  be an *regulated* probability distribution support on  $[x(E_N), x(0)]$ ,

$$\nu_{b_N}(x) := \frac{\rho_N(E(x))}{-y'(E(x))N} \circ D_{1/N} = \frac{\rho_N(E(x/N))}{-y'(E(x/N))N^2}, 0 < x(E_N) \leq x \leq x(0),$$

$$\text{where } \rho_N(E) := \frac{2^{S_0}}{2\pi^2} \sinh \left( 2\pi \sqrt{\frac{E(E_N - E)}{E_N}} \right), \quad 0 \leq E \leq E_N \quad (3.6)$$

and where  $D_{1/N}$  is the dilation operation defined in (2.46),  $E_N$  is determined via the constraint  $\int_0^{E_N} dE \rho_N(E) = N$ . As  $N \rightarrow \infty$ ,<sup>44</sup> we have the *weak convergence* of the distributions,<sup>45</sup>

$$\rho_N \rightarrow \rho, \quad N\nu_{b_N} \circ D_N \rightarrow \nu_b. \quad (3.7)$$

The partition function  $Z_n$  is evaluated in the large  $N$  limit, which poses some difficulties for us to treat it probabilistically. We thus work with the regulated probability distribution  $\nu_{b_N}$  and take the large  $N$  limit in the end. We think of this procedure as *undoing* the large  $N$  limit. Though we try to be concrete in (3.6), we shall see that the details of the regularization is not relevant for us, and any other regularization scheme can do equally well. The only requirement we demand is that the moments should converge to the moments of  $\nu_b$  large  $N$  limit. That is, for

$$m_n(\nu_{b_N}) := \int_{x(E_N)}^{x(0)} \nu_{b_N}(x) x^n dx, \quad (3.8)$$

we have

$$(Z_n/Z_1^n) m_n(\nu_b) = \lim_{N \rightarrow \infty} m_n(N\nu_{b_N} \circ D_N) = \lim_{N \rightarrow \infty} N^{1-n} m_n(\nu_{b_N}). \quad (3.9)$$

which follows from (3.7) for our regulated  $\nu_{b_N}$ .

Recall now the replica trick GPI (1.8),

$$\begin{aligned} \tilde{Z}_n &= \sum_{\pi \in \text{NC}_n} \left( \prod_{V \in \pi} Z_{|V|} \cdot \prod_{\bar{V} \in \bar{\pi}} k^{|\bar{V}|} \text{Tr} R^{|\bar{V}|} \right) \\ &= k \sum_{\pi \in \text{NC}_n} \left( \prod_{V \in \pi} k^{(|V|-1)} m_{|V|}(\nu_b) Z_1^{|V|} \cdot \prod_{\bar{V} \in \bar{\pi}} \frac{1}{k} \text{Tr} (kR)^{|\bar{V}|} \right). \end{aligned} \quad (3.10)$$

The moments  $\frac{1}{k} \text{Tr} (kR)^n$  of the given bulk radiation density matrix  $R$  define a probability distribution  $\nu_r$ . Denote the spectral values of  $R$  as  $\{\lambda_i\}_{i=1}^k$ , we have

$$\nu_r(x) := \frac{1}{k} \sum_{i=1}^k \delta(x - k\lambda_i). \quad (3.11)$$

Let's use the moments  $\tilde{Z}_n/kZ_1^n$  to define a probability distribution  $\nu_{\tilde{r}}$ , which we will soon confirm that it's indeed a valid probability distribution. The normalization  $kZ_1^n$  is to impose the normalization

<sup>44</sup>This is also known as the double scaling limit in JT gravity [70].

<sup>45</sup>We say a sequence of distributions  $\mu_N$  on  $\mathbb{R}$  converges weakly (or in weak topology) to some limiting distribution  $\mu$ , if for any bounded continuous function  $f$ , we have  $\lim_{N \rightarrow \infty} \int f(x) \mu_N(x) dx = \int f(x) \mu(x) dx$ .

that  $m_0(\nu_{\tilde{r}}) = m_1(\nu_{\tilde{r}}) = 1$  as we want for the radiation density operator  $\tilde{r}$ , which we shall discuss later. We therefore have the moments of  $\nu_{\tilde{r}}$ ,

$$m_n(\nu_{\tilde{r}}) := \frac{\tilde{Z}_n}{kZ_1^n} = k^n \sum_{\pi \in \text{NC}_n} \left( \prod_{V \in \pi} \frac{1}{k} m_{|V|}(\nu_b) \cdot \prod_{\bar{V} \in \bar{\pi}} m_{|\bar{V}|}(\nu_r) \right). \quad (3.12)$$

We can also define the moments of the corresponding distribution  $\nu_{\tilde{r}_N}$  at finite  $N$ ,

$$m_n(\nu_{\tilde{r}_N}) := \frac{k^n}{N^n} \sum_{\pi \in \text{NC}_n} \left( \prod_{V \in \pi} \frac{N}{k} m_{|V|}(\nu_{b_N}) \cdot \prod_{\bar{V} \in \bar{\pi}} m_{|\bar{V}|}(\nu_r) \right), \quad (3.13)$$

and they satisfy

$$m_n(\nu_{\tilde{r}}) = \lim_{N \rightarrow \infty} m_n(\nu_{\tilde{r}_N}). \quad (3.14)$$

Comparing the free multiplicative convolution formula (2.39) with (3.13), we see that  $\nu_{\tilde{r}_N}$  is indeed a valid probability distribution

$$\nu_{\tilde{r}_N} = \left( \nu_r \boxtimes \mu_{\frac{N}{k}, \nu_{b_N}} \right) \circ D_{\frac{N}{k}} = \nu_r \boxtimes \left( \mu_{\frac{N}{k}, \nu_{b_N}} \circ D_{\frac{N}{k}} \right) \quad (3.15)$$

where the compound Poisson distribution  $\mu_{\frac{N}{k}, \nu_{b_N}}$  follows from the free cumulants  $(N/k)m_n(\nu_{b_N})$  in (3.13), and the factor  $(k/N)^n$  in front gives rise to a dilation  $D_{N/k}$ . In the second equality we put the dilation inside the second factor of the free convolution.<sup>46</sup>

Note that the factors  $\prod_{V \in \pi} (N/k)m_{|V|}(\nu_{b_N})$  in (3.13) are the moments of the distribution  $(N/k)\nu_{b_N}$  which integrates to  $N/k$ . Since  $(N/k)\nu_{b_N}$  has bounded support, it is the unique distribution with moments  $(N/k)m_n(\nu_{b_N})$ . It is therefore not a probability distribution either. Nevertheless, the main point here is that they constitute a legit set of cumulants of a free compound Poisson distribution.

In the large  $N$  limit,  $N \rightarrow \infty$ , the cumulants  $(k/N)^{n-1}m_n(\nu_{b_N})$  converge

$$(k/N)^{n-1}m_n(\nu_{b_N}) \rightarrow k^{n-1}m_n(\nu_b), \quad (3.16)$$

so the limiting distribution of (3.15) is well-defined. We have

$$\nu_{\tilde{r}} := \lim_{N \rightarrow \infty} \nu_{\tilde{r}_N} = \nu_r \boxtimes \left( \lim_{N \rightarrow \infty} \mu_{\frac{N}{k}, \nu_{b_N}} \circ D_{\frac{N}{k}} \right). \quad (3.17)$$

We can further denote the limiting distribution in the bracket as

$$\mu_{\nu_b} := \lim_{N \rightarrow \infty} \mu_{\frac{N}{k}, \nu_{b_N}} \circ D_{\frac{N}{k}} \quad (3.18)$$

and we obtain the result claimed in our introduction (1.9), that the physical radiation spectrum is given by the convolution of the spectral data of the gravitational sector and the matter sector respectively.

$$\nu_{\tilde{r}} = \nu_r \boxtimes \mu_{\nu_b}. \quad (3.19)$$

We should emphasize that the distribution  $\mu_{\nu_b}$  and thus the end result  $\nu_{\tilde{r}}$  are independent of the way we regularize  $\nu_b$  in  $\nu_{b_N}$ . We simply made an explicit choice in (3.6) but the details of such regularization is irrelevant.

<sup>46</sup>This is legit because  $(\nu_a \boxtimes \nu_b) \circ D_\lambda$  is the spectral distribution of  $\lambda\sqrt{ab}\sqrt{a}$  for some free random variables  $a, b$  and a scalar  $\lambda$ . Clearly, we can associate the scalar multiple to  $b$  (or  $a$ ), yielding  $\nu_a \boxtimes (\nu_b \circ D_\lambda) = (\nu_a \boxtimes \nu_b) \circ D_\lambda$ .



Now we have a justification for the replica trick as a well-posed moment problem. The  $\nu_{\tilde{r}}$  defined in (3.12) is the legit probability distribution consistent with the moments given by the normalized partition functions  $\tilde{Z}_n/kZ_1^n$  in (1.8). Furthermore,  $\mu_{\nu_b}$  has compact support because both  $\nu_r$  and  $\mu_{\nu_b}$  are compactly supported. Hence, the replica trick is mathematically justified as a soluble Hausdorff moment problem which admits the *unique* solution.

Let us also quickly mention the results for a microcanonical ensemble with a fixed energy  $E \pm \Delta E$ , for which the partition function becomes much simpler,

$$Z_n[E, \Delta E] = 2^{(1-n)S} Z_1^n[E, \Delta E] , \quad (3.20)$$

where  $S := \log(\Delta E \rho(E) e^{S_0})$  is the entropy of the microcanonical ensemble. It corresponds to the partition functions (3.13) in canonical ensemble  $\nu_{\tilde{r}_N}$  for  $N = 2^S$  and  $\nu_{b_N}$  being flat. We denote the radiation spectral distribution as  $\nu_{\tilde{r}}$  and its moments read

$$m_n(\nu_{\tilde{r}}) = (2^S/k) \sum_{\pi \in \text{NC}_n} \prod_{V \in \pi} k 2^{-S} m_{|V|}(r) . \quad (3.21)$$

The corresponding spectral distribution is

$$\nu_{\tilde{r}} = (2^S/k) \mu_{k/2^S, \nu_r} + (1 - 2^S/k) \delta_0 , \quad (3.22)$$

which is a special case of the general result (3.15) we had.

Given the spectral distribution  $\nu_{\tilde{r}}$  of the radiation density operator  $\tilde{r}$  in the fundamental (bound-ary) description, its von Neumann entropy  $S(\tilde{r})$  is given by (2.10),

$$S(\tilde{r}) = -\tau(\tilde{r} \log \tilde{r}) = - \int_{\mathbb{R}_+} dx \nu_r \boxtimes \mu_{\nu_b}(x) x \log x . \quad (3.23)$$

and the Rényi entropies are given by (2.11),

$$S_\alpha(\tilde{r}) := \frac{1}{1-\alpha} \log \tau(\tilde{r}^\alpha) = \frac{1}{1-\alpha} \log \int_{\mathbb{R}_+} dx \nu_r \boxtimes \mu_{\nu_b}(x) x^\alpha . \quad (3.24)$$

We can relate them to the (standard) von Neumann entropy of the  $k \times k$  radiation density matrix  $\tilde{R}$  using (2.12),

$$S(\tilde{R}) = S(\tilde{r}) + \log k . \quad (3.25)$$

$$S_\alpha(\tilde{R}) = S_\alpha(\tilde{r}) + \frac{\alpha}{\alpha-1} \log k . \quad (3.26)$$

The convolution formula for entropy (3.23) supersedes the island formula and provides a more accurate account of the radiation entropy in the PSSY model. It applies to any entanglement spectrum of the bulk state. See Fig. 4 for an example. The Page curve can be obtained once we are given a sequence of input spectral distributions of the bulk state at different times.

Using the free multiplicative convolution, we can now resolve the spectrum even when the QES is not defined. Note that the generalized entropy is no longer relevant in our convolution formula (3.23). Islands and QES appear when the convolution factorizes. This happens when a particular summand dominates in the convolution expansion (3.12), yielding a *factorized* expression for the moments, and then the generalized entropy would show up in the entropy calculations. Such special cases hide away the more fundamental fact that the partition functions are described by a free multiplicative convolution between the quantum information encoded in the two competing QES. We will discuss when does the convolution factorize in Section 3.5.1.

### 3.2 Evaluating the free multiplicative convolution formula

We now want to evaluate free multiplicative convolution (3.17) via the  $S$ -transforms following Algorithm 2. As mentioned, this algorithm is tricky to implement for two arbitrary probability distributions. Fortunately for us, we can reformulate the calculation in terms of the  $R$ -transforms, which are better understood and easier to handle for the free compound Poisson distributions that we are dealing with. Therefore, translating everything to the  $R$ -transforms is the key to our derivation below.

We are eventually interested in the spectral distribution of (3.17), but let's first examine the case at finite  $N$  (3.15).

$$\nu_r \boxtimes \left( \mu_{\frac{N}{k}, \nu_{b_N}} \circ D_{\frac{N}{k}} \right) =: \nu_r \boxtimes \mu_{\nu_b; N} \quad (3.27)$$

where we denote the distribution in the bracket as  $\mu_{\nu_b; N}$ . We proceed by computing the  $S$ -transforms  $S_{\nu_r}$  and  $S_{\mu_{\nu_b; N}}$ . Compare the formal series of the moment function (2.35) and the  $R$ -transform (2.29), we have

$$\omega_{\nu_r}(1/\mathfrak{z}) = \mathfrak{z} R_{\mu_{1, \nu_r}}(\mathfrak{z}) \quad (3.28)$$

where we use the fact that the free compound Poisson distribution has the moments of  $\nu_r$  as cumulants  $\kappa_n(\mu_{1, \nu_r}) = m_n(\nu_r)$ . Let  $x := \omega_{\nu_r}(1/\mathfrak{z})$ , we have

$$S_{\nu_r}(x) = \mathfrak{z}(x)(1+x)/x \quad (3.29)$$

where  $\mathfrak{z}(x)$  is defined via

$$\mathfrak{z}(x) R_{\mu_{1, \nu_r}}(\mathfrak{z}(x)) = x. \quad (3.30)$$

Now consider the  $S$ -transform of  $S_{\mu_{\nu_b; N}}$ . Using (2.37), we have

$$S_{\mu_{\nu_b; N}}(x) = z(x)/x \quad (3.31)$$

where  $z(x)$  is defined via

$$z(x) R_{\mu_{\nu_b; N}}(z(x)) = x. \quad (3.32)$$

Then it follows that

$$S_{\nu_r \boxtimes \mu_{\nu_b; N}}(x) = S_{\nu_r}(x) S_{\mu_{\nu_b; N}}(x) \implies \frac{1+x}{xy(x)} = \frac{z(x)}{x} \frac{1+x}{x} \mathfrak{z}(x) \implies y(x) \mathfrak{z}(x) z(x) = x. \quad (3.33)$$

We would like to solve for the inverse function  $x(y)$ , so to obtain the Cauchy transform via (2.35) defined for any  $y \in \mathbb{C}^+$ .

$$G_{\mu_{\nu_r \boxtimes \mu_{\nu_b; N}}}(y) = y^{-1} + x(y)/y = y^{-1} + z(y) \tilde{z}(y), \quad (3.34)$$

where we've used (3.33), and the functions  $z(y)$ ,  $\tilde{z}(y)$  are given by

$$\begin{cases} \mathfrak{z} R_{\mu_{1, \nu_r}}(\mathfrak{z}) = x = yz\mathfrak{z} \\ z R_{\mu_{\nu_b; N}}(z) = x = yz\mathfrak{z} \end{cases} \implies \begin{cases} R_{\mu_{1, \nu_r}}(\mathfrak{z}(y)) = yz(y) \\ R_{\mu_{\nu_b; N}}(z(y)) = y\mathfrak{z}(y) \end{cases}. \quad (3.35)$$

Now we have a system of equations only in terms of the  $R$ -transforms. We can solve for  $\mathfrak{z}(y)$  by plugging the first equation to the second and eliminate  $z(y)$ ,

$$R_{\mu_{\nu_b; N}}[R_{\mu_{1, \nu_r}}(\mathfrak{z}(y))/y] = y\mathfrak{z}(y). \quad (3.36)$$

Since both  $\mu_{1,\nu_r}$  and  $\mu_{\nu_b;N}$  are free compound Poisson distributions and their  $R$ -transforms admit the simple form of the Lévy-Khintchine formula (2.72),

$$R_{\mu_{1,\nu_r}}(z) = \int_{\mathbb{R}_+} \frac{x}{1-xz} \nu_r(x) dx, \\ R_{\mu_{\nu_b;N}}(z) = \int_{\mathbb{R}_+} \frac{(N/k)x}{1-xz} \nu_{b_N} \circ D_{\frac{N}{k}}(x) dx = \int_{\mathbb{R}_+} \frac{(N/k)^2 x}{1-xz} \nu_{b_N}(Nx/k) dx. \quad (3.37)$$

Plugging these formulas to (3.36), we have for any  $y \in \mathbb{C}^+$ ,

$$\int_{\mathbb{R}_+} \frac{(N/k)^2 x \nu_{b_N}(Nx/k)}{y-x \int_{\mathbb{R}_+} \frac{x'}{1-x'\mathfrak{z}(y)} \nu_r(x') dx'} dx = \mathfrak{z}(y). \quad (3.38)$$

Now taking  $N \rightarrow \infty$ , we have the following equation that just depends on the two input distributions,  $\nu_r$  and  $\nu_b$ , corresponding to the moments feed into the replica trick (1.8),

$$\int_{\mathbb{R}_+} \frac{x \nu_b(x/k)/k^2}{y-x \int_{\mathbb{R}_+} \frac{x'}{1-x'\mathfrak{z}(y)} \nu_r(x') dx'} dx \stackrel{x/k \rightarrow x}{=} \int_{\mathbb{R}_+} \frac{\nu_b(x) dx}{y/x - k \int_{\mathbb{R}_+} \frac{x'}{1-x'\mathfrak{z}(y)} \nu_r(x') dx'} = \mathfrak{z}(y). \quad (3.39)$$

Note that this is a fixed-point equation for  $\mathfrak{z}(y)$  that is numerically easy to solve via an iteration algorithm, and the inputs to the algorithm are the distributions  $\nu_r$  and  $\nu_b$  that are ready to use.

Using  $\mathfrak{z}(y)$ , the Cauchy transform is given by

$$G_{\nu_r \boxtimes \mu_{\nu_b}}(y) = y^{-1} + \mathfrak{z}(y) R_{\mu_{1,\nu_r}}(\mathfrak{z}(y))/y = \frac{1}{y} \int_{\mathbb{R}_+} \frac{\nu_r(x) dx}{1-x\mathfrak{z}(y)}. \quad (3.40)$$

The spectral distribution  $\nu_{\tilde{r}} = \nu_r \boxtimes \mu_{\nu_b}$  can then be extracted using the Stieltjes inversion formula (2.25). We thus obtain the algorithm (1.10)-(1.12) claimed in the introduction.<sup>47</sup>

We should remark that such random matrices are known as the *separable sample covariance matrices*, which have applications in multivariate statistical inference [24–27] and wireless communication [71]. They read,

$$\sqrt{A_N} X_N B_N X_N^* \sqrt{A_N}, \quad (3.41)$$

where  $X_N$  is a  $N \times M$  Ginibre matrix with variance  $1/M$ , and the  $N \times N$  matrix  $A_N$  and the  $M \times M$  matrix  $B_N$  are deterministic and positive semidefinite. They are the sample covariance matrix of  $M$  samples of  $N$  data points packaged in the matrix  $\sqrt{A_N} X_N \sqrt{B_N}$ .<sup>48</sup> We will find in the next subsection that the random matrix ensemble fundamentally describing the radiation state  $\tilde{R}$  (3.51), as inferred from the convolution formula, indeed has the form of a separable sample covariance matrix.

The limiting spectral distribution of (3.41) is given by the same algorithm we presented. This result was previously obtained in random matrix literature under various weaker assumptions on the random matrix ensemble  $X_N$  [24–27]. Here we've presented a novel and much simpler way to resolve the spectrum using free probabilistic methods.

<sup>47</sup>As commented earlier in Section 2.4.3, the free multiplicative convolution is usually tricky to solve even just numerically. Nonetheless, our problem is still tractable without resorting to the machinery of the operator-valued free probability theory and the subordination formalism.

<sup>48</sup>As compared to the sample covariance matrix (2.67) mentioned earlier (cf. Footnote 41), now we allow correlations  $B_N$  among the  $M$  samples. However, these correlation is independent from the correlation  $A_N$  among the  $N$  data points. More explicitly, consider the covariance matrix  $\mathbb{E} \mathbf{Y} \mathbf{Y}^*$  of the all the sample data in stacked in one  $NM \times 1$  vector  $\sqrt{A_N} X_N \sqrt{B_N} \rightarrow \mathbf{Y} := \text{vec}(\sqrt{A_N} X_N \sqrt{B_N})$ . Then it follows that the covariance matrix is separable,  $\mathbb{E} \mathbf{Y} \mathbf{Y}^* = A_N \otimes B_N$ , and hence the name.

### 3.3 A random matrix perspective

#### 3.3.1 A model in free random variables

We now look for a density operator  $\tilde{r}$  in some  $W^*$ -probability space that has the spectral distribution of (3.17). We start by looking for the density operators  $\tilde{r}_N$  with the the spectral distribution at finite  $N$  (3.15), and then obtain  $\tilde{r}$  as the limit.

Consider a collection of  $W^*$ -probability spaces  $(\mathcal{W}_N, \tau_N)_N$  containing the following random variables  $r_N, p_N, b_N, c \in (\mathcal{W}_N, \tau_N)$ , where  $p_N$  is a projector free from the collection of the positive self-adjoint operator  $b_N$  that has the spectral distribution of  $\nu_{b_N}$  as in (3.6).<sup>49</sup>; and  $c$  is a circular element free from all  $r_N, p_N, b_N$ .<sup>50</sup> The positive self-adjoint operator  $r_N$  live in the compression  $(p_N \mathcal{W}_N p_N, \tau_{p_N})$ , where  $\tau_{p_N} := \tau/\tau(p_N) = (N/k)\tau$ , and it is also free from  $p_N$ . It has the spectral distribution of  $\nu_{r_N}$ , which is defined as

$$\nu_{r_N} := (k/N)\nu_r \circ D_{k/N} + (1 - k/N)\delta_0 . \quad (3.42)$$

It's evident that  $\tau(b_N) = \tau(r_N) = 1$ , so they are density operators.

We seek for a positive self-adjoint operator that shares the finite- $N$ -spectral distribution (3.15). We postulate that, up to a unitary,<sup>51</sup> this operator is given by

$$\tilde{r}_N := \sqrt{r_N} c b_N c^* \sqrt{r_N} \in (\mathcal{W}_N, \tau_N) . \quad (3.43)$$

Let's check if (3.43) gives us what we want. It is positive and has unit trace,

$$\tau(\tilde{r}_N) = \tau(\sqrt{r_N} c b_N c^* \sqrt{r_N}) = \tau(r_N) \tau(cc^*) \tau(b_N) = 1 \quad (3.44)$$

where we used freeness to factorize the trace, and  $\tau(cc^*) = 1$ . So it's a legit density operator. We have  $c b_N c^* \in (\mathcal{W}_N, \tau_N)$  whose spectral distribution is governed by a free compound Poisson distribution  $\mu_{1, \nu_{b_N}}$ . It is then conjugated by  $\sqrt{r_N}$  yielding the spectral distribution,

$$\nu_{\tilde{r}_N} = \nu_{r_N} \boxtimes \mu_{1, \nu_{b_N}} = (k/N)\nu_r \boxtimes \mu_{N/k, \nu_{b_N}} \circ D_{k/N} + (1 - k/N)\delta_0 \quad (3.45)$$

whose non-zero part gives (3.15).

To extract the solely the spectral distribution (3.15), consider now the  $p_N$  that projects onto the support of  $r_N$  and  $\tilde{r}_N$ . Note that both  $r_N$  and  $\tilde{r}_N$  live in the subalgebra  $(p_N \mathcal{W}_N p_N, \tau_{p_N})$ . We have  $\tau_{p_N}(r_N^n) = m_n(\nu_r)$ . The projected  $p \tilde{r}_N p$  reads

$$p_N \tilde{r}_N p_N = \sqrt{r_N} p_N c b_N c^* p_N \sqrt{r_N} \in (p_N \mathcal{W}_N p_N, \tau_{p_N}) \quad (3.46)$$

Since the projection  $p_N$  acts as a free compression on  $c b_N c^*$ , we have  $p_N c b_N c^* p_N \in (p_N \mathcal{W}_N p_N, \tau_{p_N})$  whose spectral distribution is  $\mu_{N/k, \nu_{b_N}} \circ D_{k/N}$ . This follows from (2.45). Then

$$\nu_{p_N \tilde{r}_N p_N}^p = \nu_r \boxtimes \mu_{N/k, \nu_{b_N}} \circ D_{k/N} = \nu_{\tilde{r}_N} \quad (3.47)$$

<sup>49</sup>Note that we cannot assign  $m_n(\nu_b)$  as moments to any random variable because  $\nu_b$  is not a probability distribution.

<sup>50</sup>As discussed earlier in the end of Section 2.3, the desired  $W^*$ -probability space can always be constructed to accommodate these free random variables. We can actually be more concrete by realizing  $(\mathcal{W}_N, \tau_N)$  as the free product of two Type I matrix algebras representing the gravitational sector and the matter sector respectively,  $(\mathcal{W}_N, \tau_N) = M_N(\mathbb{C}) * M_k(\mathbb{C})$ . The former hosts the observables of the JT + EOW brane system at finite  $N$  and the latter hosts the observables of the  $k$ -dimensional radiation reservoir.

<sup>51</sup>We can only determine the density operator up to a unitary transformation, because the replica trick only provides the information about the spectrum.

which matches (3.15) as we want.

We now take the large  $N$  limit. Let  $\tilde{r} \in (\mathcal{W}, \tau)$  be a random variable that has the spectrum  $\nu_{\tilde{r}}$  (3.17). Consider the limit of the sequence of the freely compressed variables  $(p_N \tilde{r}_N p_N)_N \in (p_N \mathcal{W}_N p_N, \tau_{p_N})_N \subset (\mathcal{W}_N, \tau_N)$ .<sup>52</sup> Then it follows from (3.17) that  $p_N \tilde{r}_N p_N$  converges to  $\tilde{r}$  in distribution,<sup>53</sup>

$$p_N \tilde{r}_N p_N \xrightarrow{N \rightarrow \infty} \tilde{r} . \quad (3.48)$$

Let's check the trace of  $p_N \tilde{r}_N p_N$ ,

$$\begin{aligned} \tau_{p_N}(p_N \tilde{r}_N p_N) &= \tau_{p_N}(\sqrt{r} p_N c b_N c^* p_N \sqrt{r}) = \tau_{p_N}(r) \tau_p(p_N c b_N c^* p_N) \\ &= \tau(p_N c b_N c^* p_N) / \tau(p_N) = \tau(cc^*) \tau(b_N) = 1 \end{aligned} \quad (3.49)$$

where we switched to the algebra  $(\mathcal{W}_N, \tau_N)$  in the third equality and used freeness several times. It follows from the convergence that  $\tilde{r}$  also has unit trace, so this is the density operator we want.

Let us now examine the complementary black hole system. Since the global state on the black hole-radiation is chosen to be pure, the black hole density operator  $\tilde{b}_N$  shares the same spectrum as the radiation system. We can flip the the operator ordering in (3.43) and posit that  $\tilde{b}_N$  is given by

$$\tilde{b}_N := \sqrt{b_N} c^* r_N c \sqrt{b_N} \in (\mathcal{W}_N, \tau_N) , \quad (3.50)$$

which shares the spectral distribution (3.45) as  $\tilde{r}_N$ . We should think of  $\tilde{b}_N$  as the density operator for the black hole at finite  $N$ . However, the large  $N$  limit makes it unclear if one can assign a density operator  $\tilde{b}$  to the black hole that shares this spectrum. This is unlike how we get  $\tilde{r}$  which is defined with the help of a sequence of free compressions. In fact, we claim no sensible density operator exists for the black hole at the large  $N$  limit.

Our claim is supported by the fact that in the large  $N$  limit, the distribution (3.45) has unbounded higher moments  $n > 1$ , indicating that there is not a meaningful density operator that the sequence of  $(\tilde{b}_N)_N$  converges to. Similarly, we have  $\nu_{b_N} \rightarrow \delta_0$  at the large  $N$  limit, so there isn't a density operator that  $(b_N)_N$  converges to either.

If it is the case that  $b_N$  and  $\tilde{b}_N$  do belong to the algebra of boundary observables at finite  $N$ , these limiting traces being either infinity or zero indicates that the large  $N$  algebra of boundary observables is no longer finite, such as a Type III von Neumann algebra or a Type  $\text{II}_\infty$  von Neumann algebra, so it ceases to qualify as a  $W^*$ -probability space.

Nonetheless, physically we still expect that the entropy of black hole should be identical to the radiation entropy for a global pure state. Then how can we compute the black hole entropy directly from  $\tilde{b}_N$ ? We shall provide an answer later in Section 3.4.

### 3.3.2 A random matrix model

Physically, the radiation density operators  $\tilde{r}, p_N \tilde{r}_N p_N$ , etc should be modelled as density matrices, for the radiation reservoir is simply a  $k$ -dimensional system. They are, however, fundamentally described as *random density matrices* following the intimate connection between free probability and random matrix theory. This supports the ensemble interpretation of the replica wormholes [6, 72–75].

On the other hand, the black hole system can be approximately described in a boundary random matrix theory [70]. Therefore, we can also meaningfully assign (random) density operators to model

<sup>52</sup>This is only possible for a  $W^*$ -probability space if  $\mathcal{W}$  is of Type  $\text{II}_1$  because it has no minimal projection.

<sup>53</sup>Convergence in distribution means all the moments converge. See Definition 6 for the precise definition.

the random variables  $b_N, \tilde{b}_N$  at least for finite  $N$ . In the large  $N$  limit, the operator algebra describing the boundary theory is presumably no longer a finite von Neumann algebra, so we lose the trace to define density operators. Therefore, let us stay at finite  $N$  and discuss the random matrix model that describes the outcomes of the replica trick calculation.

We shall assume that  $N$  is large enough  $N \gg k$  to approximate any large  $N$  limit result. For instance, the value of  $S_{\text{BH}}$  obtained from  $B_N$  is arbitrarily close to (3.4). At finite  $N$ , we can also embed the given  $k$ -dimensional radiation density matrix  $R$  in  $M_N(\mathbb{C})$ , so we can discuss the black hole and radiation density matrices on equal footing.

The following random matrix ensemble models (3.43),<sup>54</sup>

$$\tilde{R}_N := \sqrt{R_N} C_N B_N C_N^* \sqrt{R_N} \in (L^{\infty-}(\Omega) \otimes M_N(\mathbb{C}), \mathbb{E} \otimes \text{tr}) \quad (3.51)$$

where  $\tilde{R}_N$  is the  $N \times N$  boundary radiation density matrix;  $R_N \in M_N(\mathbb{C})$  is the embedded version of  $R \in M_k(\mathbb{C})$  in an arbitrary  $k$ -dimensional subspace of  $\mathcal{H}$  that represents  $M_N(\mathbb{C})$ ;  $B_N$  is a  $N \times N$  diagonal random matrix (in an arbitrary basis) with the spectral distribution  $\nu_{b_N}$ <sup>55</sup>; and  $C_N$  is a  $N \times N$  Ginibre matrix with variance 1. As we've alluded to, the ensemble  $R_N$  or  $P_N \tilde{R}_N P_N$  has the form of a separable sample covariance matrix (3.41).

Similarly, the random matrix model of the black hole state at finite  $N$  (3.50) reads<sup>56</sup>

$$\tilde{B}_N := \sqrt{B_N} C_N R_N C_N^* \sqrt{B_N} \in (L^{\infty-}(\Omega) \otimes M_N(\mathbb{C}), \mathbb{E} \otimes \text{tr}) . \quad (3.52)$$

The radiation density matrix at the large  $N$  limit is modelled by a  $k \times k$  density matrix under the rank- $k$  projection  $P_N$  onto the support of  $R_N$ . It is given by

$$\tilde{R} := \lim_{N \rightarrow \infty} P_N \tilde{R}_N P_N \in (L^{\infty-}(\Omega) \otimes M_k(\mathbb{C}), \mathbb{E} \otimes \text{tr}). \quad (3.53)$$

We claim that  $\tilde{R}_N, \tilde{B}_N, \tilde{R}$  model in finite dimensions the density operators  $\tilde{r}_N, \tilde{b}_N, \tilde{r}$  as defined in (3.43), (3.50), (3.48), in the sense that the spectral distributions of the former is well approximated by the spectral distributions of the latter at large  $k$  and  $N$ . We picked the normalization that  $\mathbb{E} \tilde{B}_N, \tilde{R}_N$  and  $\tilde{R}$  are all normalized with the standard trace ( $\text{Tr}$ ) to a good approximation at large  $k$  and  $N$ , so they can be viewed as density matrices.<sup>57</sup> We shall provide the justification in Appendix B.

The entropies of  $\tilde{R}, \tilde{R}_N, \tilde{B}_N$  are then also random variables on the underlying probability space  $(\Omega, \mathcal{F}, P)$ . For a separable sample covariance matrix, it is known that the almost surely spectrum of the random matrix converges (weakly) to the freely convoluted distribution  $\nu_{\tilde{r}}$  [24, 26, 27]. Namely, a typical draw of say  $\tilde{R}$  can have its spectral distribution arbitrarily close to  $\nu_{\tilde{r}}$  with high probability at large enough  $k$  and  $N$ . Therefore, we shall use  $S(\tilde{R})$  to denote both its expectation value and its *typical* entropy value.

<sup>54</sup> Note that  $L^{\infty-}(\Omega) := \bigcap_{p=1}^{\infty} L^p(\Omega)$  denotes the algebra of classical random variables that have all its moments bounded. It is larger than  $L^{\infty}(\Omega)$  and in fact not a von Neumann algebra. We need to work with it instead of  $L^{\infty}(\Omega)$  because the Ginibre matrix has Gaussian entries which have unbounded support but bounded moments. Therefore  $(L^{\infty-}(\Omega) \otimes M_N(\mathbb{C}), \mathbb{E} \otimes \text{tr})$  is only a  $*$ -probability space but not a  $W^*$ -probability space. In the end of the day, since we only demand the convergence in distribution, it's legit that the limit is described by a finite von Neumann algebra albeit the random matrices are not. (cf. Appendix B).

<sup>55</sup> More precisely, we say demand the empirical spectral distribution (cf. Appendix B) of  $B_N$  to be  $\nu_{b_N}$ . Here the basis can be chosen arbitrarily because the Ginibre ensemble is both left- and right-unitarily invariant.

<sup>56</sup> Here we treat the Ginibre matrix  $C_N$  to be independent from the ones used for (3.43). We do so because the replica trick GPI does not tell us how  $\tilde{R}$  and  $\tilde{R}_N$  are correlated, so we will not care about their potential correlations.

<sup>57</sup> Note that  $\text{tr} B_N = m_1(\nu_{b_N}) = 1$ , so unlike  $R_N, B_N$  is not a density matrix in the usual normalization convention.

When  $R$  is maximally mixed,  $R_N = P_N/k$ , so we have

$$\tilde{R} := \lim_{N \rightarrow \infty} X_N B_N X_N^* \quad (3.54)$$

where  $X_N = P_N C_N/k$  is a  $k \times N$  Ginibre ensemble with variance  $1/k$ . It matches with the random matrix model proposed by PSSY in their Appendix D (eqn. D.10). In this view, the PSSY model with non-flat bulk entanglement spectrum simply amounts to conjugating the random matrix with the additional matrix  $R$ .

Let us put the random matrix ensemble in a more familiar form in which a matrix model is usually written. According to PSSY, we can rewrite the large  $N$  limit of  $B_N$  in terms of the random JT Hamiltonian  $H_N$  found by [70] in the double scaling limit as follows,

$$Z_n/Z_1^n = m_n(\nu_b) = \lim_{N \rightarrow \infty} N^{1-n} \mathbb{E} \text{tr} B_N^n = \lim_{N \rightarrow \infty} \int dH_N (\text{Tr } x(H_N)^n) e^{-N \text{Tr} V_N(H_N)} \quad (3.55)$$

where we used (3.8) and  $x(\cdot)$  is the spectral function  $y(E)$  in (3.1) rescaled by  $1/Z_1$ , and the integral is over  $N \times N$  Hermitian matrices  $H_N$  with the measure controlled by the a sequence of potentials  $(V_N(H_N))_N$  that are determined by  $(\nu_{b_N})_N$ .<sup>58</sup> We can then compute any spectral function  $\mathbb{E}f(H)$  of the JT Hamiltonian via

$$\mathbb{E}f(H) = \lim_{N \rightarrow \infty} \int dH_N f(H) e^{-N \text{Tr} V_N(H_N)}. \quad (3.56)$$

On the other hand, we can write the expectation taking over the Ginibre ensemble as

$$\mathbb{E}_{C_N}(\cdot) = \pi^{-N^2} \int dC_N(\cdot) e^{-\text{Tr} C_N C_N^*}. \quad (3.57)$$

For a rectangular  $k \times N$  Ginibre ensemble  $X_N$ , we have

$$\mathbb{E}_{X_N}(\cdot) = \pi^{-Nk} \int dX_N(\cdot) e^{-\text{Tr} X_N X_N^*}. \quad (3.58)$$

Say we are interested any spectral function  $f(\tilde{R})$ , such as the von Neumann entropy. Putting things together, it can be evaluated using the following matrix integral,

$$\mathbb{E}f(\tilde{R}) = \lim_{N \rightarrow \infty} \pi^{-Nk} \int dX_N dH_N f(R X_N x(H_N) X_N^* R) e^{-N \text{Tr} V_N(H_N) - \text{Tr} X_N X_N^*} \quad (3.59)$$

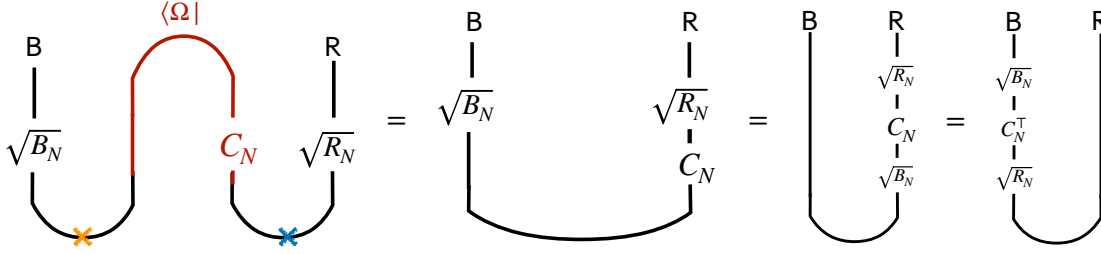
where  $R$  is the given deterministic  $k \times k$  bulk radiation density matrix.

### 3.3.3 A generalized Page's model

The above random matrix models can also be packaged together in a random tensor network. We now take a slight detour to introduce some basics of RTN. Generally, a RTN is used as a model for capturing the information-theoretic aspects of the AdS/CFT duality, and it is very successful in reproducing features like the Ryu-Takayanagi formula and the subregion-subregion duality [16, 35, 76–79].

A RTN is defined for a given a connected undirected graph  $(V, E)$ . For each edge-vertex pair  $(e, x)$ , we assign a Hilbert space  $\mathcal{H}_{e,x}$ . Attached to each edge  $e = (xy) \in E$  is a bipartite system  $\mathcal{H}_e := \mathcal{H}_{e,x} \otimes \mathcal{H}_{e,y}$  with dimension  $D_e$ , which we refer as the bond dimension. An entangled pure state  $|\phi_e\rangle \in \mathcal{H}_e$  is assigned to each edge  $e \in E$ . Usually, they are chosen as Bell states. The vertices are

<sup>58</sup>This sequence diverges in accordance with the fact that  $(\nu_{b_N})_N$  approaches zero.



**Figure 5: The PSSY model as a generalized Page's model.** The left tensor diagram shows the RTN for the PSSY model with non-flat entanglement spectrum, where the red part indicates the projection to a Gaussian random vector. The middle diagram shows that it generalizes Page's model by allowing the two edges to have non-flat spectrum. One can shift the operators all to the radiation system or the black hole system, so when tracing out either  $B$  or  $R$ , we obtain the consistent random matrix models deduced from our free convolution formula. The two phases described by island formula corresponds to make the blue cut or the orange cut which entail an entropy contribution of  $S_{\text{BH}}$  or  $S(R)$  respectively.

divided into bulk vertices  $V_b$  and boundary vertices  $V_\partial$ . At each bulk vertex, we introduce a projection to a random Gaussian vector<sup>59</sup>  $|\psi_x\rangle \in \bigotimes_{e \sim x} \mathcal{H}_{e,x}$  where  $e \sim x$  denote any edge  $e$  connected to  $x$ . In this way, the RTN defines a state in the boundary Hilbert space,

$$|\rho_\partial\rangle := \bigotimes_{x \in V_b} \bigotimes_{e \in E} (I_{V_\partial} \otimes \langle \psi_x |) |\phi_e\rangle \in \bigotimes_{x \in V_\partial} \mathcal{H}_x . \quad (3.60)$$

Because of the random Gaussian projection,  $|\rho_\partial\rangle$  is only normalized in expectation with small fluctuations at large bond dimensions.

The RTN of the PSSY model consists of two edges, one bulk vertex and two boundary vertices (cf. Fig. 5). The two edges states are given by

$$|\sqrt{R_N}\rangle_{\text{RR}'} := (\sqrt{R_N})_{\text{R}} \otimes I_{\text{R}'} |\Omega\rangle_{\text{RR}'} , \quad |\sqrt{B_N}\rangle_{\text{BB}'} := (\sqrt{B_N})_{\text{B}} \otimes I_{\text{B}'} |\Omega\rangle_{\text{BB}'} \quad (3.61)$$

where the Bell state  $|\Omega\rangle$  has dimension  $N$  in both states but the Schmidt rank of them are  $|\text{R}| = k$  and  $|\text{B}| = N$  respectively. The density matrices  $B_N$  and  $R_N$  are diagonal in the basis used to define the Bell state  $\Omega$ . They represent the purifications of the effective radiation density matrix  $R_N$ , and of the effective black hole density matrix  $B_N$ . If a minimal (QES) surface cuts through one of the two bonds, it entails an entropy contribution of  $S(R_N) = S(R)$  or  $S_{\text{BH}}$  respectively.

The random projection acts on the purifying systems  $\text{R}'\text{B}'$ ,

$$|\psi\rangle_{\text{R}'\text{B}'} = (C_N)_{\text{R}'} \otimes I_{\text{B}'} |\Omega\rangle_{\text{R}'\text{B}'} \quad (3.62)$$

Then the RTN defines the state

$$(I_{\text{BR}} \otimes \langle \psi |_{\text{R}'\text{B}'} | \sqrt{R_N}\rangle_{\text{RR}'} | \sqrt{B_N}\rangle_{\text{BB}'} \quad (3.63)$$

<sup>59</sup>A random Gaussian vector has i.i.d. complex Gaussian entries with zero mean and unit variance. It is unnormalized, and it is equivalent to a Haar random vector up to the normalization. Nonetheless, the norm  $\|\psi_x\|$  is one on expectation and it is asymptotically normalized at large dimensions in the sense that the fluctuations of the norm tends to zero.



which equals to either of the following two expressions,

$$(\sqrt{R_N}C_N\sqrt{B_N})_R \otimes I_B|\Omega\rangle_{RB}, \quad I_R \otimes (\sqrt{B_N}C_N^\top\sqrt{R_N})_B|\Omega\rangle_{RB} \quad (3.64)$$

where  $C_N$  (or  $C_N^\top$ ) is a  $N \times N$  Ginibre ensemble. The equalities follow from manipulating the tensor diagrams in Fig. 5.<sup>60</sup> Let's denote its marginal states as  $\tilde{B}_N$  and  $\tilde{R}_N$  respectively as usual for the fundamental description of the respective density matrices. It is evident that they match with (3.51) and (3.52) respectively.<sup>61</sup>

This RTN is manifestly a generalization of the model proposed by Page to deduce the Page curve [10, 11]. Page considered a microcanonical scenario, where the  $B_N$  state is flat and  $\log N$  measures the entropy of the remaining black hole. Also,  $R_N$  is chosen to be flat. Then one simply has a random bipartite state on BR and Page considered the entropy  $S(\tilde{R}_N)$  as one vary the relative sizes between  $N$  and  $k$ . Basically, the RTN model we described generalizes Page's model by allowing the two bond states to be non-flat.

In this view, we should think of the two competing saddles as the competition between two “areas”, with one given by the standard BH area term and the other given by the bulk entanglement between the brane and the radiation. The island formula describes the situation in which the radiation entropy can be described by a clean cut of one of the two bonds. This happens when the contribution coming from contracting the B system is separated from R, in the sense that the resulting spectrum of  $\tilde{R}$  depends only on  $B_N$  or  $R_N$  but not both. We will later see how these separations are mathematically characterized by the one-shot entropies.

Generally, due to the non-flatness of the area spectrum, there isn't such a clean cut between the two sources of correlations. Nonetheless, the random Ginibre ensemble help set the two sources free, so we can exploit this freeness to calculate the spectrum of (3.51). This is the general takeaway message from our work. See more discussions in Section 4. To what extent freeness is relevant and how useful it is in AdS/CFT shall be explored in future work.

### 3.4 Hawking's surprisal

Hopefully we have convinced you that the free convolution formula for the radiation entropy (3.23) doesn't entail the notion of generalized entropy. Nonetheless, it can actually be interpreted as a generalized entropy if formulated in the right way. Consider the following formulation of the generalized entropy in Wall's proof of the generalized second law [29]. Wall equates the generalized entropy associated with any cut  $\gamma$  on the event horizon of a (classically) stationary black hole as follows,<sup>62</sup>

$$S_{\text{gen}}(\gamma) = A(+\infty)/4G_N - S(\psi|_{\mathcal{A}_\gamma}||\psi_{\text{HH}}|_{\mathcal{A}_\gamma}), \quad (3.65)$$

where  $\psi$  is the quantum state of the matter field in consideration,  $\psi_{\text{HH}}$  is the Hartle-Hawking state describing a black hole in thermal equilibrium, the relative entropy is evaluated for the subalgebra  $\mathcal{A}_\gamma$  associated with the spacetime subregion external to  $\gamma$ , and  $A(+\infty)$  is the horizon area at infinity. It

<sup>60</sup>When moving the operators from B to R or vice versa, they generally pick up a transpose defined w.r.t. to basis chosen in defining the Bell state. The diagonal matrices  $B_N$  and  $R_N$  are not affected, so it only affects the Ginibre matrix. Nonetheless, since the entropies of a Ginibre matrix is i.i.d.,  $C_N^\top$  is the same random ensemble as any  $N \times N$  Ginibre ensemble.

<sup>61</sup>Note that the RTN also determines how the reduced random density states are correlated, but we will ignore this aspect of the RTN model as the gravity calculation does not inform us about it. cf. Footnote 56.

<sup>62</sup>The formula is valid for a classically stationary black hole for which the contribution of the classical area increase is turned off, which is the interesting case concerning the generalized second law.

is independent of the state  $\psi$  so it drops out when we study the evolution of  $S_{\text{gen}}$  along the horizon. The relative entropy term is the important one. Its change packages together the variations of the entropy of quantum fields and their black-reactions to the horizon area, and the generalized second law simply follows from its monotonicity under restriction to subalgebras associated with nested spacetime subregions [80–82].

The idea is that although the free convolution formula for entropy (3.23) does not have the form of an area piece plus a bulk entropy piece, it could nevertheless match with Wall’s formula of the generalized entropy (3.65). This was also the idea recently used to show that in de Sitter spacetime, the entropy of any semiclassical state<sup>63</sup> is matching with the generalized entropy of the cosmological horizon [83].

We are interested in the relative entropy between the black hole state  $\tilde{b}_N$  and the black hole thermal state  $b_N$ . Since we can interpret this quantity as *Hawking’s surprisal* in learning the actual black hole state  $\tilde{b}_N$ , given his original belief that the black hole is basically a canonical thermal state  $b_N$ . First of all, recall the definition (3.50) of  $\tilde{b}_N$ ,<sup>64</sup>

$$\tilde{b}_N := \sqrt{b_N} c^* r_N c \sqrt{b_N} \in (\mathcal{W}, \tau) . \quad (3.66)$$

Technically, since we do not have appropriately defined random variables for the black hole and radiation in the asymptotic limit, we consider the relative entropy  $S(\tilde{b}_N || b_N)$  at finite  $N$ , and then take the large  $N$  limit. At finite  $N$ , Umegaki’s formula (2.17) of the relative entropy applies as  $\tilde{b}_N$  and  $b_N$  are density operators in a  $W^*$ -probability space  $(\mathcal{W}, \tau)$ ,

$$S(\tilde{b}_N || b_N) = \tau(\tilde{b}_N \log \tilde{b}_N) - \tau(\tilde{b}_N \log b_N) . \quad (3.67)$$

The first term gives

$$\tau(\tilde{b}_N \log \tilde{b}_N) = -S(\tilde{b}_N) = -S(\tilde{r}_N) \quad (3.68)$$

where we used the fact that the  $\tilde{b}_N$  shares the spectral distribution with  $\tilde{r}_N$ .

The second term gives

$$\begin{aligned} \tau(\tilde{b}_N \log b_N) &= \lim_{n \rightarrow 0} \tau(\tilde{b}_N b_N^n) / n = \lim_{n \rightarrow 0} \tau(c^* r_N c b_N^{n+1}) / n \\ &= \tau(r_N) \tau(cc^*) \lim_{n \rightarrow 0} \tau(b_N^{n+1}) / n = S(b_N) \end{aligned} \quad (3.69)$$

where we used the replica trick for the logarithm, substituted in (3.50), and used freeness to compute the traces; in the last step we used  $\tau(r_N) = 1, \tau(cc^*) = 1$ .

Together we have,

$$S(\tilde{b}_N || b_N) = S(b_N) - S(\tilde{r}_N) = (S(b_N) + \log N) - (S(\tilde{r}_N) + \log N) . \quad (3.70)$$

where we added a pair of  $\pm \log N$  to each term to prepare for taking the large  $N$  limit.

Now we take the large  $N$  limit. The second term in bracket gives,

$$\lim_{N \rightarrow \infty} S(\tilde{r}_N) + \log N = S(\tilde{r}) + \log N = S(\tilde{R}) , \quad (3.71)$$

<sup>63</sup>One can meaningfully assign an entropy to states because the operator algebra of observables in de Sitter is shown to be finite, more specifically, a Type II<sub>1</sub> factor.

<sup>64</sup>Unlike entropies that are spectral functions, the relative entropy not only depends on spectra of  $\tilde{b}_N$  and  $b_N$ , but also the relative phases between them. Hence, we have made a choice in defining  $\tilde{b}_N$  to be (3.50) that fixes the relative phase between  $\tilde{b}_N$  and  $b_N$ .

where the last step follows from (2.12).

The first term in bracket gives,

$$\begin{aligned}
\lim_{N \rightarrow \infty} S(b_N) + \log N &= - \lim_{N \rightarrow \infty} \int dx \nu_{b_N}(x) x \log x + \log N \\
&= - \lim_{N \rightarrow \infty} \int N^{-1} dx (N^2 \nu_{b_N}(x)) (x/N) \log(x/N) \\
&\stackrel{y:=x/N}{=} - \lim_{N \rightarrow \infty} \int dy (N^2 \nu_{b_N}(yN)) y \log y \\
&\stackrel{(3.7)}{=} - \int dy \nu_b(y) y \log y = S_{\text{BH}} ,
\end{aligned} \tag{3.72}$$

which is nothing but the BH entropy (3.4).

Altogether we have

$$\lim_{N \rightarrow \infty} -S(\tilde{b}_N || b_N) = S(\tilde{R}) - S_{\text{BH}} . \tag{3.73}$$

The minus relative entropy on the LHS could be interpreted as the black hole/radiation entropy measured with reference to the black hole background, and the RHS says that it simply equals to the radiation entropy  $S(\tilde{R})$  with the BH entropy subtracted.

Put (3.73) differently, we have

$$S(\tilde{R}) = S_{\text{BH}} - \lim_{N \rightarrow \infty} S(\tilde{b}_N || b_N) , \tag{3.74}$$

which is now a formula for the radiation entropy that doesn't explicitly involve free probability. Unlike the free convolution formula (3.23), (3.74) manifestly obeys the Bekenstein bound [84, 85], i.e.  $S(\tilde{R}) \leq S_{\text{BH}}$ , also known as the central dogma of black hole physics [8]. This follows from the positivity of the relative entropy.

This formula addresses the question we left in section 3.3. How one can make sense of the black hole entropy at the large  $N$  limit in terms of  $\tilde{b}_N$  which doesn't have a sensible large  $N$  limit? The answer is that we should simply compute Hawking's surprisal. Although both  $\tilde{b}_N$  and  $b_N$  at large  $N$  are no longer legit random variables in a finite von Neumann algebra, the limiting relative entropy,  $\lim_{N \rightarrow \infty} S(\tilde{b}_N || b_N)$ , survives and remains finite.

The limiting relative entropy already resembles the term in Wall's formula, where the reference state is also the thermal state of the black hole. In order to make the match more manifest, we can try to be more explicit about the limiting relative entropy with a formal construction. Let the GNS Hilbert space constructed from the tracial state  $\tau$  be  $\tilde{\mathcal{H}}_{\text{TFD}}$ . Any faithful normal states admit a representation as a cyclic and separating state vector  $|\mathbf{1}\rangle$ . The state vector for the state  $x \mapsto \tau(b_N x)$  is

$$|\psi_N\rangle := |\sqrt{b_N}\rangle := \sqrt{b_N} |\mathbf{1}\rangle , \tag{3.75}$$

which is a finite  $N$  version of the thermofield double state (i.e. the canonical purification of the thermal state). We can obtain other states from the action of the operators in  $\mathcal{W}$  or its commutant  $\mathcal{W}' \subset \mathcal{B}(\tilde{\mathcal{H}}_{\text{TFD}})$  on  $|\sqrt{b_N}\rangle$ ,

$$a |\psi_N\rangle = |a \sqrt{b_N}\rangle, \quad a \in \mathcal{W} ; \quad a' |\psi_N\rangle = |\sqrt{b_N} a'\rangle, \quad a' \in \mathcal{W}' . \tag{3.76}$$

Then various inner products are given by

$$\langle \psi_N | b^* a | \psi_N \rangle = \psi_N(b^* a) = \tau(b^* a b_N), \quad a, b \in \mathcal{W} \tag{3.77}$$

$$\langle \psi_N | a'^* a | \psi_N \rangle = \tau(a'^* \sqrt{b_N} a \sqrt{b_N}), \quad a \in \mathcal{W}, a' \in \mathcal{W}' \quad (3.78)$$

Another the state vector we care about in  $\tilde{\mathcal{H}}_{\text{TFD}}$  is the one for  $\tilde{b}_N$ ,

$$|\tilde{\psi}_N\rangle := \sqrt{r_N} c |\psi_N\rangle = |\sqrt{b_N} c \sqrt{r_N}\rangle, \quad r_N, c \in \mathcal{W}' . \quad (3.79)$$

Then we can write the relative entropy equivalently using Araki's formula (2.16),

$$S(\tilde{b}_N || b_N) = -\langle \tilde{\psi}_N | \log \Delta_{\tilde{\psi}_N | \psi_N} | \tilde{\psi}_N \rangle . \quad (3.80)$$

where  $\Delta_{\tilde{\psi}_N | \psi_N}$  is the relative modular operator.

Now we follow a standard procedure to define the Hilbert space  $\mathcal{H}_{\text{TFD}}$  at the large  $N$  limit. A pedagogical description is given in Section 3 of [86]. The basic idea is to apply the GNS construction for the state  $\psi_{\text{HH}}$  obtained from the limits of the finite  $N$  counterparts  $\psi_N$ . In particular the inner product between  $a|\psi_{\text{HH}}\rangle$  and  $b|\psi_{\text{HH}}\rangle$  are obtained from the limit of the corresponding finite  $N$  inner products  $\langle \psi_N | b^* a | \psi_N \rangle = \psi_N(b^* a)$ .  $\psi_{\text{HH}}$  is called the thermofield double state, also known as the Hartle-Hawking (HH) state for a thermal black hole. Then we complete the algebra  $\mathcal{W}$  to a von Neumann algebra  $\mathcal{A}$  acting on  $\mathcal{H}_{\text{TFD}}$  in the weak topology. Unlike  $\mathcal{W}$ , the so obtained large  $N$  algebra  $\mathcal{A}$  is generically infinite.

We can also repeat the procedure to define the large  $N$  limit of  $|\tilde{\psi}_N\rangle$ . The constructed GNS Hilbert space would be the same as  $\mathcal{H}_{\text{TFD}}$ , as long as for any  $a \in \mathcal{W}$  there exists some  $b \in \mathcal{W}$  such that the inner product  $\langle a\psi_N | b\tilde{\psi}_N \rangle$  has a finite large  $N$  limit. We shall assume that this is the case.

Since all the inner products in  $\mathcal{H}_{\text{TFD}}$  are defined via the limits, we have

$$\lim_{N \rightarrow \infty} -\langle \tilde{\psi}_N | \log \Delta_{\tilde{\psi}_N | \psi_N} | \tilde{\psi}_N \rangle = -\langle \tilde{\psi} | \log \Delta_{\tilde{\psi} | \psi_{\text{HH}}} | \tilde{\psi} \rangle (= S(\tilde{\psi} || \psi_{\text{HH}})) , \quad (3.81)$$

so formally, under the assumption that we could put both  $\tilde{\psi}$  and  $\psi_{\text{HH}}$  in the Hilbert space  $\mathcal{H}_{\text{TFD}}$ , we could rewrite (3.74) as

$$S(\tilde{R}) = S_{\text{BH}} - S(\tilde{\psi} || \psi_{\text{HH}}) . \quad (3.82)$$

In the PSSY model, the JT black hole is stationary so  $S_{\text{BH}}$  can be matched with the constant term  $A(+\infty)/4G_N$  in (3.65). Therefore, our formula (3.82), or the less explicit (3.74), matches with Wall's formula (3.65) and thus can be viewed as a generalised entropy. Just like (3.65), the more relevant is the relative entropy term if we only care about the evolution of the radiation entropy, as in the Page curve for example. Here the relative entropy term contains implicitly the free convolution formula in its first argument.

For more realistic evaporating black holes, matter fields back-react on the geometry creating area fluctuations. This is not captured by the PSSY model, and would make the GPI analysis much harder. Therefore, being able to evaluate the exact radiation entropy of a realistic evaporating black hole using free multiplicative convolution is perhaps too good to hope for. The upshot of the relative entropy formula (3.82) is that it might be applicable in more general settings.

### 3.5 Sanity checks

We now perform some sanity checks to our result to show its consistency with some other relevant results in the literature.

### 3.5.1 The island formula and the one-shot entropies

When is the island formula (1.5) valid? The general answer was proposed by AP, who claim that the QES prescription in AdS/CFT is only valid in regimes governed by the *one-shot entropies* [16]. These entropies characterize information processing tasks in the one-shot regime where the only one copy of the source is involved. This is in contrast to how the Shannon/von Neumann entropy is relevant in the asymptotic regime where many identical and independent copies of the source are assumed.

In the context of the island formula in the PSSY model, the relevance of these one-shot entropies was first discussed in [17]. The main point is that the island formula only holds when a black hole is still young such that the *smooth min-entropy* of its emitted radiation (in the effectively description) is below  $S_{\text{BH}} - \mathcal{O}(\sqrt{\beta})$ , or when a black hole gets old enough such that the *smooth max-entropy* of emitted radiation (in the effectively description) is above  $S_{\text{BH}} + \mathcal{O}(\sqrt{\beta})$ .

We have shown that the free convolution formula offers an accurate value of  $S(\tilde{R})$  in any regime. Therefore, the convolution formula should reduce to the island formula under the above one-shot entropic conditions. We have observed this phenomenon in Fig. 4 where the transitions occur at  $S_{\text{BH}} = 15$ , corresponding to the min-entropy, and  $S_{\text{BH}} = 35$ , corresponding to the max-entropy. These transition are further smoothed out by the finite temperature corrections. Our goal in this subsection is to show that the convolution factorizes precisely under the above mentioned entropic conditions. Instead of directly comparing the size of each summand to single out the dominant term, we resort to the generalized Page’s model introduced previously that realizes this spectral distribution.<sup>65</sup>

We first introduce the one-shot entropies, which for us specifically refer to the smooth min-entropy and the smooth max-entropy.<sup>66</sup>

$$S_{\text{max}}^\varepsilon(A) = \min_{A' \in \mathcal{B}_A^\varepsilon} \log \text{rank}(A'), \quad S_{\text{min}}^\varepsilon(A)_\rho = \max_{A' \in \mathcal{B}_A^\varepsilon} -\log \lambda_{\text{max}}(A'). \quad (3.83)$$

where  $\mathcal{B}_A^\varepsilon$  denotes the  $\varepsilon$ -ball around a density matrix  $A$  in the purified distance.<sup>67</sup> The non-smoothed min/max-entropies are the two extreme points of Rényi  $\alpha$ -entropies for  $\alpha = \infty$  and  $\alpha = 0$  respectively. For an  $N \times N$  density matrix  $A$ , both  $S_{\text{max}}^\varepsilon(A)$  and  $S_{\text{min}}^\varepsilon(A)$  are upper bounded by  $\log N$ . Generally, one needs to deal with the conditional smooth entropies, but the setup in the PSSY model is relatively simply in that there are no quantum side information to condition on. So the above defined one-shot entropies will suffice for our purpose.

The operational relevance of these one-shot entropies was first pointed out in [88].<sup>68</sup> The smooth min-entropy measures the maximum amount of randomness, which is  $\varepsilon$ -close to uniform, that can be extracted from a single copy of an  $N \times N$  density matrix  $A$ . More concretely, a random projection  $\Pi_{\text{min}}$  of dimension smaller  $2^{S_{\text{min}}^\varepsilon(A)}$  can typically produce a uniform outcome, i.e.  $N\Pi_{\text{min}}A\Pi_{\text{min}} \approx \Pi_{\text{min}}$ ,<sup>69</sup> hence extracting  $S_{\text{min}}^\varepsilon(A)$  bits of perfect randomness from a non-perfect source of  $\log N$  bits.

<sup>65</sup>This strategy is the same as how AP argued the relevance of the one-shot entropies in the corrected QES prescription [16] using RTN and the decoupling theorems, whereas a direct comparison among the saddle contributions is carried out in a replica trick GPI derivation in [17].

<sup>66</sup>The max-entropy is usually defined with the Rényi-1/2 entropy rather than the Rényi-0 (Hartley) entropy that we used, as the former is preferred for the duality relations that are particularly useful when dealing with the conditional versions. We stick to the latter for its better intuition as measuring the rank of the state. In any case, their difference is often small after smoothing. cf. Tomamichel [87] for a comprehensive review.

<sup>67</sup>The purified distance is a fidelity-based metric between density matrices defined as  $\delta(A, B) := \sqrt{1 - F(A, B)}$  where the fidelity is defined as  $F(A, B) := (\text{Tr} \sqrt{\sqrt{A}B\sqrt{A}})^2$ . For our purposes, the trace distance would work just fine as well.

<sup>68</sup>See the more comprehensive discussion including the conditional one-shot entropies in [89].

<sup>69</sup>It also follows from the special case of a one-shot decoupling theorem (cf. Lemma 4.5 in [90]) applied to a scenario without the reference system.

On the other hand, we can faithfully compress a copy of the  $N$ -dimensional state  $A$  into a “smaller” state with dimension  $2^{S_{\max}^\epsilon(A)}$  up to an error of  $\mathcal{O}(\epsilon)$ . More explicitly, we can project the state to the spectral support on the largest  $2^{S_{\max}^\epsilon}$  eigenvalues, such that the projected state  $\Pi_{\max} A \Pi_{\max}$  (with appropriate renormalization) remains  $\epsilon$ -close to  $A$ . Note that such a projector  $\Pi_{\max}$  commutes with  $A$ . Since the projected operator has its trace  $\epsilon$ -close to 1, we shall neglect the normalization issue here.

Consider now the random density matrix for the radiation (3.51). When  $S_{\min}^\epsilon(B_N) - 1 > S_{\max}^\epsilon(R)$ , for two independently chosen smoothing parameters  $\epsilon$  and  $\epsilon$ , we can find a projector  $\Pi$  with rank  $M$  such that  $S_{\min}^\epsilon(B_N) - 1 > M > S_{\max}^\epsilon(R)$  and

$$\Pi R_N \Pi = \sqrt{R_N} \Pi \sqrt{R_N} \approx R_N \quad (3.84)$$

which follows from  $M > S_{\max}^\epsilon(R)$ . Note that  $R \in (M_k(\mathbb{C}), \text{Tr})$  and  $R_N \in (M_N(\mathbb{C}), \text{Tr})$  share the same values for any entropy. It follows that

$$\begin{aligned} \tilde{R}_N &= \sqrt{R_N} C_N B_N C_N^* \sqrt{R_N} \approx \Pi \sqrt{R_N} \Pi C_N B_N C_N^* \Pi \sqrt{R_N} \Pi \\ &\approx \Pi \sqrt{R_N} \Pi U |C_N| B_N |C_N| U^* \Pi \sqrt{R_N} \Pi \approx \Pi \sqrt{R_N} \Pi \sqrt{R_N} \Pi \approx R_N \end{aligned} \quad (3.85)$$

where we replace each  $R_N$  by  $\Pi \sqrt{R_N} \Pi$  and use the polar decomposition  $C_N = U |C_N|$  with a Haar random unitary  $U$  (cf.(2.59)). Then we have  $\Pi U |C_N| B_N |C_N| U^* \Pi \approx \Pi$  under a random projection  $\Pi U$  when  $S_{\min}^\epsilon(B_N) - 1$ . To see how the condition follows, recall that the min-entropy depends only on the largest eigenvalue. Since  $|C_N|^2/N = C_N C_N^*/N$  asymptotes to a squared quarter-circular element has its spectral support over  $[0, 2]$ , the min-entropy  $S_{\min}(|C_N| B_N |C_N|/N)^{70}$  is only different from  $S_{\min}(B_N)$  by one bit, and so is the smooth min-entropy close by continuity. This explains the subtraction by one that we need to include in the condition above.<sup>71</sup>

Therefore, we have

$$S(\tilde{R}) \approx S(\tilde{R}_N) \approx S(R_N) = S(R) , \quad (3.86)$$

so a young black hole satisfying the condition  $S_{\min}^\epsilon(B_N) - 1 > S_{\max}^\epsilon(R)$  obeys the island formula (1.5). Generally, the approximation error that (3.85) entails is set by the smoothing parameter  $\mathcal{O}(\epsilon)$ .

When  $S_{\max}^\epsilon(B_N) < S_{\min}^\epsilon(R)$ , we simply invert the argument above and apply it to (3.52). There exists another projector  $\Pi$  with rank  $M$  such that  $S_{\max}^\epsilon(B_N) < M < S_{\min}^\epsilon(R) - 1$  and

$$\Pi B_N \Pi = \sqrt{B_N} \Pi \sqrt{B_N} \approx B_N , \quad (3.87)$$

which follows from  $M > S_{\max}^\epsilon(B_N)$ . It follows that

$$\begin{aligned} \tilde{B}_N &\approx \Pi \sqrt{B_N} \Pi C_N R_N C_N^* \Pi \sqrt{B_N} \Pi \\ &\approx \Pi \sqrt{B_N} \Pi U |C_N| R_N |C_N| U^* \Pi \sqrt{B_N} \Pi \approx \Pi \sqrt{B_N} \Pi \sqrt{B_N} \Pi \approx B_N . \end{aligned} \quad (3.88)$$

Therefore, we have

$$S(\tilde{R}) \approx S(\tilde{R}_N) = S(\tilde{B}_N) \approx S(B_N) , \quad (3.89)$$

<sup>70</sup>The positive operator  $|C_N| B_N |C_N|/N$  is normalized to a good approximation at large  $k$ , so we can treat it as a density matrix.

<sup>71</sup>This order-one effect is somewhat universal. It already appears in Page’s original calculation as the leading correction term [10].

so an old black hole satisfying the condition  $S_{\max}^\epsilon(B_N) < M < S_{\min}^\epsilon(R) - 1$  obeys the island formula (1.5). Again, the approximation error that (3.88) entails is set by the smoothing parameter  $\mathcal{O}(\varepsilon, \epsilon)$ .

This is the general result addressing when one can compute the entropy of a boundary reduced state by putting a minimal cut as in Fig. 5 in the generalized Page’s model, without invoking the free convolution formula. While we assume the density matrix  $R$  to be arbitrary, we do know more about  $B_N$  which is only parameterized by the temperature (assuming the brane tension is large  $\mu \gg 1/\beta$ ). In fact,  $B_N$  essentially behaves like a projector of dimension  $2^{S_{\text{BH}}}$  in the following sense. In the PSSY model, the relevant parameter is  $\beta$ . The smooth min/max-entropies of  $B_N$  are roughly

$$S_{\min}^\epsilon(B_N) \approx S_{\text{BH}} - \mathcal{O}(1/\sqrt{\beta}), \quad S_{\max}^\epsilon(B_N) \approx S_{\text{BH}} + \mathcal{O}(1/\sqrt{\beta}). \quad (3.90)$$

for  $\epsilon \sim \mathcal{O}(\text{poly}(\beta))$ . This nearly flatness is in fact a generic feature of a holographic state, such as a thermal black hole state. Its smooth min/max-entropy is close to its von Neumann entropy, which is of order  $1/G_N$ , up to a subleading  $1/\sqrt{G_N}$  correction [91]. In the PSSY model, the temperature plays the parametric role of  $G_N$ , and with dimensional analysis one can artificially introduce Newton’s constant as  $\beta \rightarrow \beta G_N$ .

We thereby establish our claim made in the beginning of this subsection. The value of the smoothing parameters  $\varepsilon$  and  $\epsilon$  can be chosen freely so that one obtain the most useful estimate of the entropy  $S(\tilde{R})$ . For  $B_N$ , a good choice is something like  $\epsilon \sim \mathcal{O}(\text{poly}(\beta))$  that gives the estimate (3.90) up to an error of  $\mathcal{O}(-\log \beta)$ . Such finite temperature effect that smoothes out the Page transition is well understood [6, 92–94]. The correction is of order  $\mathcal{O}(1/\sqrt{\beta})$ , which is a subleading effect as compared to the case of leading corrections  $\sim \mathcal{O}(1/\beta)$  when the gap between the min-entropy and the max-entropy of the bulk radiation state is large as considered in [16, 17].

While this finite temperature correction is usually treated in parallel to the corrections due to non-flat bulk spectrum, here we see that it’s really the same phenomenon of enhanced replica symmetry breaking due to the non-flatness. We treated them on equal footing because a RTN models them equally. It’s just that the non-flatness of the thermal spectrum is mild so it only gives a subleading correction to the island formula, whereas the spectrum of the bulk state could set made arbitrary by quantum processing.

### 3.5.2 The result of PSSY

As a second sanity check, we now reproduce the result concerning the radiation spectrum in the original work of PSSY. They considered a maximally entangled bulk state.  $\nu_r = \delta_1$  and  $m_n(\nu_r) = 1, \forall n$ , so the operator  $r$  is the identity. Then the radiation spectrum is given by the gravitational sector only

$$\nu_{\tilde{r}} = \mu_{\nu_b} = \lim_{N \rightarrow \infty} \mu_{\frac{N}{k}, \nu_{b_N}} \circ D_{\frac{N}{k}}, \quad (3.91)$$

which is simply a compound Poisson distribution up to a rescaling. Its  $R$ -transform admits the general form of the Lévy-Khintchine formula (2.72),

$$R_{\mu_{\nu_b}; N}(z) = (N/k) \sum_{n=0}^{\infty} z^n \int_{\mathbb{R}_+} (N/k) x^{n+1} \nu_{b_N}(Nx/k) dx = \int_{\mathbb{R}_+} \frac{(N/k)^2 x}{1-xz} \nu_{b_N}(Nx/k) dx. \quad (3.92)$$

The large  $N$  limit gives,

$$R_{\mu_{\nu_b}}(z) = \int_{\mathbb{R}_+} \frac{x/k^2}{1-xz} \nu_b(x/k) dx \stackrel{x/k \rightarrow x}{=} \int_{\mathbb{R}_+} \frac{x}{1-kxz} \nu_b(x) dx. \quad (3.93)$$

To extract the distribution, we resort to the Cauchy transform via (2.31),

$$G_{\mu_{\nu_b}}(z)R_{\mu_{\nu_b}}(G_{\mu_{\nu_b}}(z)) = \int_{\mathbb{R}_+} \frac{xG_{\mu_{\nu_b}}(z)}{1 - kxG_{\mu_{\nu_b}}(z)} \nu_b(x) dx = zG_{\mu_{\nu_b}}(z) - 1, \quad (3.94)$$

which, up to the normalization convention used, matches with PSSY's equation (2.34) that is derived from the Schwinger-Dyson equations for the resolvents.

To extract the spectral distribution, one needs to solve for the Cauchy transform. However, even in this case, a closed-form expression for the spectral distribution is not possible. PSSY offered an approximated solution and we now try to make a similar approximation by deriving it directly from our general convolution formula.

The approximation starts by working with the regulated probability distributions. Let us set  $N = k$  in (3.13) as a first approximation,

$$\begin{aligned} m_n(\mu_{\tilde{r}_k}) &= \sum_{\pi \in \text{NC}_n} \left( \prod_{V \in \pi} m_{|V|}(\nu_{b_k}) \cdot \prod_{\bar{V} \in \bar{\pi}} m_{|\bar{V}|}(\nu_r) \right) = \sum_{\pi \in \text{NC}_n} \prod_{V \in \pi} m_{|V|}(\nu_{b_k}) \\ &= m_n(\nu_{b_k}) = \int_{x(E_k)}^{x(0)} dx x^n k^{n-1} \frac{\rho_k(E(x))}{-y'(E(x))} \end{aligned} \quad (3.95)$$

where we used a simple approximation in the second line that  $\pi = \eta$  dominates the sum. This is because  $|\pi| \leq |\eta| = n$  and  $n \mapsto m_n(\nu_{b_k})$  is increasing for large  $k$ . In the last step, we used (3.6) and  $E_k$  is defined via

$$\int_0^{E_k} dE \rho_k(E) = k. \quad (3.96)$$

In order to make an explicit comparison with PSSY's result of the radiation spectrum, we need to switch to their normalization convention,

$$\int D(x) dx = k, \quad \text{Tr} \tilde{R}^n = \int D(x) x^n dx. \quad (3.97)$$

where  $D(x)$  denote the spectral distribution of the radiation state in their convention.

On the other hand, we have

$$k^{-1} \text{Tr} \left( k \tilde{R} \right)^n \approx m_n(\mu_{\tilde{r}_k}) \implies \text{Tr} \tilde{R}^n = \int_{x(E_k)}^{x(0)} dx x^n \frac{\rho_k(E(x))}{-y'(E(x))}, \quad (3.98)$$

so it follows that

$$D(x) = \frac{\rho_k(E(x))}{-y'(E(x))} = \int_0^{E_k} dE \rho_k(E) \delta(x - x(E)) \quad (3.99)$$

where we used the properties of delta functions to bring it to a form that matches with PSSY's equation (2.50) where  $x(E) = y(E)/Z_1$  is denoted as  $\omega(E)$  in PSSY.

We remark that we used a different regularization scheme from PSSY in making the approximation. Instead of using a smooth integrable distribution  $\rho_k$  as in (3.6), they simply truncate  $\rho_E$  at  $E'_k$  when  $\int_0^{E'_k} \rho(E) = k$  and replaces the contribution from  $E \in (E'_k, +\infty)$  by a delta function that gives a lowest possible spectral value  $\lambda_0$ . In our scheme,  $\lambda_0$  is replaced by  $x(E_k)$ . Nonetheless, the regularization schemes only differ at large  $E$  for which  $x(E)$  is in a small exponential tail. The qualitative result should not depend on the exact scheme used in making the approximation.



### 3.5.3 The result of Cheng et al on random tensor networks

We mentioned in our introduction that a similar result was obtained recently in the context of random tensor network by Cheng et al [35]. They considered a generalized RTN model where the edges are no longer maximally entangled but rather having arbitrary entanglement spectrum, modeling the area fluctuations that one would have in AdS/CFT. In the case of there are two competing minimal cuts  $\gamma_{1,2}$ , the transition between them is no longer sharp. The details of the entropy of the reduced state on  $\Gamma \subset V_{\partial}$  depend on the spectra  $\mu_{1,2}$  of the bonds cut through by  $\gamma_{1,2}$ .

This is essentially the same phenomenon that we are investigating for the PSSY model, as was made obvious in the previous subsection. There is, however, one conceptual difference between our starting points and it is a somewhat pedantic. The RTN is by construction finite dimensional and the above mentioned result is formulated for a sequence of RTNs with increasing bond dimensions and converging spectral distributions at the two cuts. The free probability description is thus invoked via Voiculescu's theorem. We, however, do not have the convenience of a random matrix/tensor model as the fundamental description. Therefore, we take the first-principle rationale that GPI in the planar limit is the fundamental description, and then we do not need to worry about the convergence issues. Technically, this means we shall understand the GPI in terms of free random variables rather than random matrices.

One can certainly question the validity of the planar limit, and indeed it is, after all, a compromise for the analytic tractability of the problem, because we do not have a controlled way to account for the small fluctuations due to the nonplanar diagrams. So we opt for the pragmatic approach to work in the planar limit, and we think it's conceptually cleaner, because we can translate the entire GPI calculation into the framework of non-commutative probability theory.

Alternatively, if one takes the random matrix model, which can be deduced from the relevant free random variables, as the fundamental description for the PSSY model, one has essentially the same setup as the RTN model in Cheng et al. Then the free probabilistic calculation only serves as a leading order approximation. Heuristically, how well it approximates a finite dimensional RTN also gauges how well justified the planar approximation is.

Besides that, our result in Section 3.1 and their results in Section 3.2 are almost identical. At the first sight, it might not be obvious that the results are actually identical. The main technical difference comes from that we are dealing with the thermal distribution at the large  $N$  limit, which is itself not a proper probability distribution. Nonetheless, if we focus on our result at finite  $N$  (3.15), then a comparison can be made. Consider the operator we had (3.43), whose non-zero part gives (3.15),

$$\sqrt{r_N} c b_N c^* \sqrt{r_N} . \quad (3.100)$$

We can reorder the product without changing its moments into the following product of two free factors

$$r_N \cdot c b_N c^* \quad (3.101)$$

where  $r_N$  has the spectral distribution  $\nu_{r_N}$  (3.42).

The second term shares the spectrum with  $b_N c^* c = b_N q^2$ , where we again have a product of two free factors. Recall that a squared quarter-circular element  $q^2$  has the distribution (2.56), which we denoted as MP(1) following Cheng et al.  $b_N$  has the spectral distribution  $\nu_{b_N}$  (3.6). Then the moments of (3.100) are shared with the moments of

$$r_N \cdot b_N \cdot q^2 , \quad (3.102)$$

and its moments define the distribution

$$\nu_{r_N} \boxtimes \nu_{b_N} \boxtimes \text{MP}(1) , \tag{3.103}$$

which matches exactly with their Theorem 3.4 in [35].

## 4 Takeaway: Freeness is good

We presented a concrete solution to the PSSY model with non-flat bulk entanglement spectrum using ideas from free probability theory. We showed that the free probabilistic framework is perhaps the most appropriate language to understand the model, bridging between the replica trick GPI calculation on one hand and the random matrix model on the other.

Aside from the solution to this particular problem, the main conceptual takeaway from our work is that freeness is a useful feature that can be exploited in studying how gravity processes quantum information, in parallel to the more familiar notion of independence. Though certainly more subtle and underlying than tensor independence, it is worth the attention. By making the PSSY model a bit more complicated, we are able to see freeness manifest in gravity that is otherwise hidden when we only consider a flat entanglement spectrum. As we mentioned, this has to do with a mathematical fact that the identity is trivially free from any other random variable, and this is the only situation where freeness and independence coincides. Therefore, a flat entanglement spectrum with the reduced density operator being an identity hides the fact that the degrees of freedom associated with two candidate surfaces in transition are not always independent but rather always free.

Generally, it is hard to characterize quantum correlation in gravity just like in any generic many-body quantum system. We usually rely on the independence between subsystems in order to make definitive statements. For instance, the reference that encodes any information that falls into a black hole is *decoupled*, i.e. independent from, the radiation or the remaining black hole before or after the Page time respectively. This allows us to tell where is the information encoded so we can in principle retrieve it from there.

Over the past, we have made much progress in understanding how gravity stores and processes quantum information, starting from the seminal work of Page [10, 11] and Hayden-Preskill [95] towards the more recent quantum error correction with complementary recovery [96–98], (random) tensor networks [76, 78, 91, 99–101] and entanglement wedge reconstruction as state-merging [16]. From an information-theoretic perspective, an overarching technical core underlying these results is the decoupling theorems [90, 102–104], which characterizes when do we have independence between certain subsystems. Then Uhlmann’s theorem [105] would tell us how the quantum correlations is constrained so that we can characterize the whereabouts of the information. Another example is the quantum de Finetti theorem [106–109], which uses the permutation symmetry of large collection of systems to deduce the independence among subsystems. This idea is recently used to explain the tension between the recent replica trick calculation of the Page curve and Hawking’s original calculation [110].

On the other hand, as we have emphasized in Section 2.3, freeness is a distinct notion of independence. In contrast to the standard tensor independence, that concerns commuting random variables, freeness is a notion of independence tailor-made for non-commuting random variables. Technically, it is defined via the free product instead of the tensor product. As for the decoupling theorem, de Finetti theorem and Uhlmann’s theorem, freeness comes with its own tricks that we can utilize. There is a whole package of free probabilistic tools that one can use to quantitatively characterize the information, as we have seen them in action.

More importantly, freeness and independence are not mutually exclusive in applications as they work under different conditions and concern different objects. The work by Cheng et al is another good example of this. While freeness can be used to give precise results when the spectral distributions at the candidate minimal cuts converge, one can still make some useful claims about the boundary reduced state using the decoupling approach when the convergence is not assumed (cf. Section 4 in [35]).

From the RTN perspective, the intuition is that the random tensors not only facilitate the decoupling among different bulk subregions, but also set the information encoded in different QES free from each other. It would be interesting to explore the possibility of putting them to work in synergy in future works.

In quantum theory, which is usually treated as a non-commutative probability theory of observables, we are used to the notion of independence as captured by commuting observables. Now we have seen that it's also possible and fruitful to accommodate freeness in quantum theory, and especially quantum gravity. We hold the vision that freeness can be very useful in quantum gravity. One hint comes from the relevance of the free independence between two perturbation of a black hole separated longer than the scrambling time [111]. We are also largely motivated by the fact that various recent GPI calculations call for random matrix models [70, 112–114]. In fact, people already came across free random variables in studying the large  $N$  matrix models [115–117]. Free probability is very relevant here because of its intimate connection with random matrix theory, building on the remarkable insight of Voiculescu that independent random matrices tend to be free in the large  $N$  limit. The theory of higher order freeness has also been developed to study the fluctuations of random matrices [118–121]. We therefore believe that freeness should be ubiquitous in gravity, provided that we look in the right directions.

## Acknowledgments

I am grateful to Roland Speicher for helping me with free probability theory, and Renato Renner for many stimulating discussions and valuable comments. I also thank Chris Akers, Patrick Hayden, Juan Maldacena, Tony Metger, Geoff Penington, Xiaoliang Qi, Pratik Rath, Stephen Shenker, Douglas Stanford, Edward Witten, and Zhenbin Yang for discussions.

## A Additivity violation of the minimum channel output entropy

In this appendix, we discuss how an old black hole encodes a channel that violates the additivity of minimum channel output entropy.

In quantum information theory, understanding the quantum channel capacity has fundamental significance. It provides us fundamental insights regarding how much advantage that quantumness can offer over classical resources. One specific problem concerns the capacity of quantum channel for transmitting classical information. People want to know if entanglement can help given the access to multiple copies of the channel. It is conjectured that quantum entanglement does not enhance the channel capacity and the classical capacity of quantum channel is additive. The additivity is verified for many concrete channels but remains an open problem until a counterexample is found.

The conjecture is not proved false directly. The counter example concerns an equivalent problem of the additivity of the minimum channel output entropy, which we shall now briefly introduce. The

minimum channel output entropy<sup>72</sup> for a channel  $\mathcal{N}$  is defined as<sup>73</sup>

$$S_{\min}(\mathcal{N}) := \min_{\psi} S(\mathcal{N}(\psi)). \quad (\text{A.1})$$

Shor showed that asking about channel additivity is equivalent to, among several other equivalent statements, asking if  $S_{\min}$  is additive for any two channels  $\mathcal{N}, \mathcal{M}$  [30].

$$S_{\min}(\mathcal{N} \otimes \mathcal{M}) \stackrel{?}{=} S_{\min}(\mathcal{N}) + S_{\min}(\mathcal{M}) \quad (\text{A.2})$$

This is proved false by Hastings [31] who showed that

$$S_{\min}(\mathcal{N} \otimes \bar{\mathcal{N}}) < S_{\min}(\mathcal{N}) + S_{\min}(\bar{\mathcal{N}}) \quad (\text{A.3})$$

for some  $N$ -dimensional channel  $\mathcal{N}$  and its conjugate channel  $\bar{\mathcal{N}}$ .<sup>74</sup>

We've been using free probability to analyze the black hole information problem, whereas randomized constructions are in fact ubiquitous in quantum information. The rationale of injecting randomness dates back to Shannon's founding work of proving the channel coding theorem using a random coding [122]. Hastings' counterexample is also a random construction, and the existence of additivity violation is argued probabilistically. He didn't find an explicit channel that gives the violation, but rather  $\mathcal{N}$  is a random unitary channel of form

$$\mathcal{N}(\cdot) := \sum_i^d p_i U_i(\cdot) U_i^* \quad (\text{A.4})$$

with independently sampled Haar random unitaries  $U_i$ , weighted by some carefully chosen probability vector  $p_i$  that is roughly uniform. Then Hastings showed that in the regime of  $1 \ll d \ll N$ , some (or in fact any typical) instance of the random channel ensemble  $\mathcal{N}$  violates additivity as in (A.3).

Hastings' work was built upon the earlier work of Hayden-Winter whose proved false the Rényi versions of the conjecture [123], where Hayden-Winter used a similar random construction in the Stinespring representation with random isometries.

There are many follow-ups to the Hastings-Hayden-Winter (HHW) breakthrough [32, 124–128], but they almost all use the same family of random channels with minor variations. Unfortunately, no concrete counterexamples nor extensive violations are found.

Let us mention a particular variant of the HHW-type construction that is been analyzed using free probability [32].<sup>75</sup> Consider the following random channel mapping from the  $N$ -dimensional quantum state space to itself,

$$\mathcal{E}(\cdot) := \frac{1}{d} \sum_i^d C_i T(\cdot) T C_i^*, \quad T := \sqrt{d} \left( \sum_{i=1}^d C_i^* C_i \right)^{-\frac{1}{2}} \quad (\text{A.5})$$

where we have  $n$  independent  $N \times N$  Ginibre matrices  $C_i$ , and  $T$  is a *rectification* operator that ensures the channel is trace-preserving. Note that  $T$  almost surely converge to identity at large  $d$ , so the rectification is a technicality that is only relevant for finite  $d$  and can be approximately ignored at large  $d$ .

<sup>72</sup>Here  $S_{\min}$  should not be confused with the min-entropy. We apologize for the clash of notations.

<sup>73</sup>This is a convex optimization problem so the inputs are restricted to pure states w.l.o.g.

<sup>74</sup>The conjugate channel of a channel  $\mathcal{N}(\cdot) := \sum_i K_i(\cdot) K_i^*$  is defined as  $\bar{\mathcal{N}}(\cdot) := \sum_i \bar{K}_i(\cdot) \bar{K}_i^*$ .

<sup>75</sup>I thank Tony Metger for pointing out this reference to me.

In the regime

$$1 \ll d \ll N, \quad (\text{A.6})$$

it's shown that typical instances of the random channel  $\mathcal{E}$  (together with its conjugate channel  $\bar{\mathcal{E}}$ ) violates the additivity with a small gap [32]. The proof is built upon the earlier work by Collins [127]. Note that (A.5) is a variant of the HHW-type counterexample because Ginibre matrices can be thought of as partially-traced Haar random unitaries/isometries (cf. [127] and the discussion in section 4.1 in [32]).

Note that the output of  $\mathcal{E}$  has a nearly flat spectrum. (cf. Theorem 3.1 in [127] for a precise statement. ) This is a universal feature of the HHW-type construction. The output of the channel is very close to a maximally mixed state of the maximal possible rank.<sup>76</sup> It's somewhat ironic that the only family of channels that we know to have superadditive capacity has almost no practical use.

Consider now the spectrum of Hawking radiation in a microcanonical ensemble for a very old black hole. Namely, we are in the regime

$$1 \ll k/2^S \ll k. \quad (\text{A.7})$$

where as assume  $k/2^S$  is an integer.

The microcanonical spectral distribution of the radiation reads (3.22),

$$\nu_{\tilde{r}} = (2^S/k)\mu_{1,\nu_r}^{\boxplus k/2^S} + (1 - 2^S/k)\delta_0. \quad (\text{A.8})$$

Now we look for a quantum channel that takes bulk radiation density operator  $R$  as input and the output has its spectrum described by  $\nu_{\tilde{r}}$  to a good approximation at large dimension  $N$ . Following the discussion in 3.3.1, a density operator with this spectrum is

$$\tilde{r} = (k/2^S)pcrc^*p \quad (\text{A.9})$$

where  $c$  is a circular element and  $p$  is a free compression of trace  $\tau(p) = 2^S/k$ . However, when we model  $\tilde{r}$  in random matrices, it doesn't take the form of a trace-preserving channel from  $r$  to  $\tilde{r}$ . We need to consider something else.

Let's restrict to the non-zero part of  $\nu_{\tilde{r}}$  and ask what channel has output spectrum described by  $\mu_{1,\nu_r}^{\boxplus k/2^S}$  up to the normalization. This is the spectrum of  $\sum_{i=1}^{k/2^S} c_i r c_i^*$  for  $k/2^S$  free circular elements  $\{c_i\}_{i \in [k/2^S]}$ . We can put in a normalization factor to obtain a completely positive trace-preserving map on  $(\mathcal{W}, \tau)$ ,

$$E(r) := \frac{2^S}{k} \sum_{i=1}^{k/2^S} c_i r c_i^*. \quad (\text{A.10})$$

Using again the connection between random matrices and free probability, we can model  $E(\cdot)$  by  $\mathcal{E}(\cdot)$  in (A.5) at large  $N$ . Then in the regime (A.7), which is identified with (A.6) through  $k/2^S = d, k = N$ , one would find that this channel violates the additivity of minimum output entropy.

Therefore, we conclude that the HHW-type random channel (A.5), which is originally constructed purely for technical convenience in approaching the additivity conjecture, can be physically encoded by an evaporating black hole at late time. However, it still begs the question what is the physical meaning

---

<sup>76</sup>It means that either the channel itself or its complement channel is close to a fully depolarizing channel. It is known that a channel is additivity-violating if and only if its complement channel is also additivity-violating [129].

of the channel  $x \mapsto E(x)$ . Could it be interpreted as a holographic map from bulk to boundary in some appropriate sense?

It's important to remark that this violation is only visible here because we sum over all possible geometries, albeit that most of them are subleading and usually deemed unimportant at late time. If we only care about entropies, the island formula is good enough as the GPI at late time is dominated by the symmetric replica wormhole saddle alone. This is in accordance with the fact that the channel  $\mathcal{E}$  is only slightly different from a depolarizing channel that outputs a maximally mixed state of rank  $d = k/2^S$ . However, it is exactly this subtle difference that is responsible for the violation in the HHW counterexamples. Therefore, in order to see the additivity violation from an old black hole, taking replica symmetry breaking into account is indeed indispensable.

It's plausible that realistic black holes at late time encode information via a more sophisticated (random) channel such that one can hope for more counterexamples to additivity or even large extensive violations. This possibility is indeed contemplated in [130] by relating it to the existence of certain entanglement properties of black hole microstates. However, as we learnt in the PSSY model, one perhaps needs to wield the full GPI in order to extract such channels, which remains technically challenging for realistic black holes.

## B More on random matrices

In this appendix, we introduce the notion of random matrices and their spectral distributions. We shall see how random matrices make contact with free probability.

### B.1 Preliminaries

We have the probability space given by a Kolmogorov triple  $(\Omega, \mathcal{F}, P)$ . The sample space  $\Omega$  is equipped with some  $\sigma$ -algebra  $\mathcal{F}$  and a probability measure  $P$ . We can simply take the  $\mathcal{F}$  to be generated by the relevant random variables. Consider the matrix elements of an  $N \times N$  matrix  $A$  as random variables. They are defined as measurable functions,  $A_{ij} : \omega \mapsto A_{ij}(\omega)$ , from  $\Omega$  to  $\mathbb{C}$  equipped with the Borel  $\sigma$ -algebra. Then we define the random matrix  $A$  as  $A : \Omega \rightarrow M_N(\mathbb{C}), \omega \mapsto A(\omega)$ . We can view it as the joint random variable of its matrix elements  $(A_{ij})_{i,j=1}^N$  mapping from  $\Omega$  to  $\mathbb{C}^{N^2}$  equipped with the Borel  $\sigma$ -algebra. We can then pushforward  $P$  to define the probability distribution of the matrix elements  $\nu(A_{ij})$  on  $\mathbb{C}$ .

We are particularly interested in the spectra of random matrices. For a  $N \times N$  matrix  $A$  with eigenvalues  $\{\lambda_i(A)\}_{i=1}^N$ , we can define its *spectral distribution*, which is a probability measure on  $\mathbb{C}$ , as an average over the Dirac mass on each eigenvalue,

$$\nu_A(\lambda) := \frac{1}{N} \sum_{i=1}^N \delta(\lambda - \lambda_i(A)) . \tag{B.1}$$

Often we need to consider a sequence of random matrices of size  $N \in \mathbb{N}$ , denoted as  $(A_N)_N$ . For technical convenience,<sup>77</sup> we assume these random matrices are all defined on the same probability space  $(\Omega, \mathcal{F}, P)$ .

Let the eigenvalues of  $A_N(\omega)$  be  $\{\lambda_i(A_N(\omega))\}_{i=1}^N$ . Then we define the random variables  $\{\lambda_i(A_N)\}_{i=1}^N$  to be measurable functions from  $\Omega$  to  $\mathbb{C}$ ,  $\lambda_i(A_N) : \omega \mapsto \lambda_i(A_N(\omega))$ . Let the set of probability measures on  $\mathbb{C}$  be  $\mathcal{M}(\mathbb{C})$ . Then we define the *empirical spectral distribution* (ESD) of  $A_N$  as a measurable

---

<sup>77</sup>Generally we need to work with the probability space consisting of the product space  $\prod_N \Omega_N$ , with the product  $\sigma$ -algebra  $\prod_N \mathcal{F}_N$  and the product measure  $\prod_N P_N$ .

function from  $\Omega$  to  $\mathcal{M}(\mathbb{C})$ ,

$$\nu_{A_N} : \omega \mapsto \nu_{A_N(\omega)}(\lambda) := \frac{1}{N} \sum_{i=1}^N \delta(\lambda - \lambda_i(A_N(\omega))), \quad (\text{B.2})$$

which is a *probability measure-valued* random variable.<sup>78</sup>

For a random variable, we can discuss different notions of convergence as we send  $N \rightarrow \infty$ . In particular, we need the notion of almost sure convergence. For a sequence of classical (scalar) random variables  $(X_N)_N$  on  $\mathbb{C}$ , it converges almost surely to some random variable  $X$  if

$$P\left(\omega \in \Omega : \lim_{N \rightarrow \infty} X_N(\omega) = X(\omega)\right) = 1. \quad (\text{B.3})$$

Let us restrict to Hermitian matrices. For a sequence of ESD  $(\nu_{A_N})_N$  of hermitian matrices  $(A_N)_N$ , we are interested in the cases that the sequence converges to a deterministic limit.

The strongest convergence is the almost sure convergence. We say a sequence of ESD  $(\nu_{A_N})_N$  converges almost surely to a measure  $\nu_a \in \mathcal{M}(\mathbb{R})$  if for any bounded continuous function  $f : \mathbb{R} \mapsto \mathbb{R}$ ,  $\int \nu_{A_N}(\lambda) f(\lambda) d\lambda$ , which is a scalar random variable, converges almost surely to the value of  $\int \nu_a(\lambda) f(\lambda) d\lambda$ . In other words, we almost surely have the ESD  $(\nu_{A_N})_N$  converge to  $\nu_a$  in *weak topology*. Then we call  $\nu_a$  the *limiting spectral distribution* (LSD) of the random matrix ensemble  $(A_N)_N$ .

For classical random variables, the almost sure convergence implies convergence in probability, which in turn implies convergence in distribution.<sup>79</sup> For a sequence of ESD, the convergence in probability is defined similarly as the almost sure convergence up to changing the sense of convergence, whereas the convergence in distribution should correspond to the notion of *convergence in expectation*. It concerns the *expected ESD*  $\mathbb{E}\nu_{A_N}$  defined as

$$\int \mathbb{E}\nu_{A_N}(\lambda) f(\lambda) d\lambda := \mathbb{E} \int \nu_{A_N}(\lambda) f(\lambda) d\lambda \quad (\text{B.4})$$

for all bounded continuous function  $f : \mathbb{R} \mapsto \mathbb{R}$ . Then we say  $(\nu_{A_N})_N$  converges in expectation to some probability measure  $\nu_a$  on  $\mathbb{R}$  if the sequence of *deterministic* distributions  $(\mathbb{E}\nu_{A_N})_N$  converges weakly to  $\nu_a$ , i.e.

$$\lim_{N \rightarrow \infty} \int \mathbb{E}\nu_{A_N}(\lambda) f(\lambda) d\lambda = \int \nu_a(\lambda) f(\lambda) d\lambda \quad (\text{B.5})$$

for all bounded continuous function  $f : \mathbb{R} \mapsto \mathbb{R}$ . Of course, almost sure convergence implies convergence in expectation. In summary, for a sequence of ESDs, almost sure convergence implies convergence in probability which implies convergence in expectation.

## B.2 Making contact with free probability

To see how asymptotic limit of large random matrices make contact with free probability, we would like to abstract away the Kolmogorov triple and resort to the non-commutative probability space

<sup>78</sup>The notion of an ESD is perhaps confusing at the first encounter, essentially because we have two types of randomness put together. It would be conceptually more transparent when we work with a non-commutative probability space in section 2.

<sup>79</sup>For a sequence of classical (scalar) random variables  $(X_N)_N$  on  $\mathbb{R}$ , it converges in probability if for any  $\varepsilon$ ,  $\lim_{N \rightarrow \infty} P(|X_N(\omega) - X(\omega)| > \varepsilon) = 0$ ; and it converges in distribution if  $\lim_{N \rightarrow \infty} \mathbb{E}f(X_N) = \mathbb{E}f(X)$  for any bounded continuous function  $f : \mathbb{R} \mapsto \mathbb{R}$ , i.e. the sequence of distributions  $\nu_{X_N}$  converges weakly to  $\mu_X$ .

$(L^{\infty-}(\Omega) \otimes M_N(\mathbb{C}), \mathbb{E} \otimes \text{tr})$ , which is a  $*$ -probability space (cf. Footnote 54). We now discuss the convergence of random matrices in terms of the convergence of non-commutative random variables.

Note that because ESD itself is a probability measure-valued random variable the convergence of ESD entails two notions convergences. For example, the almost sure convergence we've been discussing concerns the random variable itself, whereas we demand each instance to converge weakly as a probability measure. The latter is also known as the convergence in distribution for the random variables that carry these instances of probability measures. Since we've stashed the Kolmogorov triple and opted for  $(L^{\infty-}(\Omega) \otimes M_N(\mathbb{C}), \mathbb{E} \otimes \text{tr})$  to describe random matrices, we take the expectation and hence left the deterministic distributions  $\mathbb{E}\nu_{A_N}$ , the only notion of convergence available is the convergence in expectation, or convergence in distribution for the respective random variables in  $L^{\infty-}(\Omega) \otimes M_N(\mathbb{C})$ .

**Definition 6** (Convergence (in distribution) of non-commutative random variables). *Let  $(\mathcal{A}_N, \tau_N)_N$  be a sequence of non-commutative probability spaces, and let  $(\mathcal{A}, \tau)$  be another non-commutative probability space. Let  $(a_{N;1})_N, \dots, (a_{N;d})_N \in (\mathcal{A}_N, \tau_N)_N$  be a collection sequences of random variables, and  $a_1, \dots, a_d \in (\mathcal{A}, \tau)$ , we say  $(a_{N;1})_N, \dots, (a_{N;d})_N$  converges in distribution to  $a_1, \dots, a_d$  if*

$$\lim_{N \rightarrow \infty} \tau_N(a_{N;i_1} \dots a_{N;i_k}) = \tau(a_{i_1}, \dots, a_{i_k}) \quad (\text{B.6})$$

for any  $\{i_1, \dots, i_k\} \in [d]$ . We denote it as

$$(a_{N;i})_N \rightarrow a_i, \quad i \in [d]. \quad (\text{B.7})$$

Concretely, for a sequence of random matrices  $(A_N)_N \in (\mathcal{A}_N, \tau_N)_N$ , we can now talk about its limit at  $N \rightarrow \infty$  as some non-commutative random variable  $a$  in  $(\mathcal{A}, \tau)$  as long as the moments match asymptotically. Equivalently, when the ESD  $(\nu_{A_N})_N$  converge in expectation to  $\nu_a$ , we can identify  $\nu_a$  as the (deterministic) spectral distribution for some  $a \in (\mathcal{A}, \tau)$ .

With the appropriate notion of convergence, we can now discuss the when do large random matrices become free from each other.

**Definition 7** (Asymptotic freeness). *A collection of sequences of non-commutative random variables  $(a_{N;1})_N, \dots, (a_{N;d})_N \in (\mathcal{A}_N, \tau_N)_N$  is asymptotically free if they converge in distribution to some elements  $a_1, \dots, a_d$  respectively in some non-commutative probability space  $(\mathcal{A}, \tau)$  such that the collection  $a_1, \dots, a_d$  is free.*

When we choose  $(\mathcal{A}_N, \tau_N)_N$  to be  $(L^{\infty-}(\Omega) \otimes M_N(\mathbb{C}), \mathbb{E} \otimes \text{tr})_N$ , asymptotic freeness implies that the ESD  $\nu_{A_N}$  converge in expectation to  $\nu_a$ . Hence, asymptotic freeness for random matrices is a property defined at the level of expectations, capturing a relatively weak sense of asymptotic freeness. More explicitly, consider (2.18) for any two sequences of random matrices  $(A_N)_N$  and  $(B_N)_N$ . If for any finite collection of integers  $\{m_i, n_i\}_i \geq 0$  such that,

$$\lim_{N \rightarrow \infty} \frac{1}{N} \mathbb{E} \text{Tr} \left( \prod_i \left( (A_N)^{m_i} - \frac{1}{N} \mathbb{E}[\text{Tr}(A_N)^{m_i}] \cdot I_N \right) \cdot \left( (B_N)^{n_i} - \frac{1}{N} \mathbb{E}[\text{Tr}(B_N)^{n_i}] \cdot I_N \right) \right) = 0, \quad (\text{B.8})$$

then  $(A_N)_N$  and  $(B_N)_N$  are asymptotically free.

Asymptotic freeness does not only concern random matrices, due to the flexibility in choosing the non-commutative probability spaces  $(\mathcal{A}_N, \tau_N)_N$ . While one can choose  $(\mathcal{A}_N, \tau_N)_N$  to be  $(L^{\infty-}(\Omega) \otimes M_N(\mathbb{C}), \mathbb{E} \otimes \text{tr})_N$  for  $N \times N$  random matrices, we can also discuss asymptotic freeness for sequences of deterministic matrices in  $(M_N(\mathbb{C}), \text{tr})_N$  when they converge in distribution to elements in  $(\mathcal{A}, \tau)$ . We can exploit this to define the stronger notion of *almost sure asymptotic freeness* for random matrices.



**Definition 8** (Almost sure asymptotic freeness). *A collection of sequences of random  $N \times N$  matrices  $(A_{N;1})_N, \dots, (A_{N;d})_N$  in  $(L^{\infty-}(\Omega) \otimes M_N(\mathbb{C}), \mathbb{E} \otimes \text{tr})_N$  is almost surely asymptotically free if, for almost all  $\omega \in \Omega$ , the sequences  $(A_{N;1}(\omega))_N, \dots, (A_{N;d}(\omega))_N \in (M_N(\mathbb{C}), \text{tr})_N$  are asymptotically free, i.e., they converge in distribution to some elements  $a_1, \dots, a_d$  respectively in some non-commutative probability space  $(\mathcal{A}, \tau)$  such that the collection  $a_1, \dots, a_d$  is free.*

Note that here we demand the elements in  $(M_N(\mathbb{C}), \text{tr})$  under fixed  $\omega \in \Omega$  to converge in distribution for almost all  $\omega$ , which implies the almost sure convergence of the ESDs. Therefore, almost sure asymptotic freeness is stronger than asymptotic freeness.

More explicitly, consider again (2.18) for any two sequences of random matrices  $(A_N)_N$  and  $(B_N)_N$  with each admitting an asymptotic spectral distribution  $\nu_a$  and  $\nu_b$  respectively. If for any finite collection of integers  $\{m_i, n_i\}_i \geq 0$  such that the following random variable converges almost surely

$$\left( \frac{1}{N} \text{Tr} \prod_i ((A_N)^{m_i} - \nu_a(\lambda^{m_i}) \cdot I_N) \cdot ((B_N)^{n_i} - \nu_b(\lambda^{n_i}) \cdot I_N) \right)_N \xrightarrow{a.s.} 0, \quad (\text{B.9})$$

then  $(A_N)_N$  and  $(B_N)_N$  are almost surely asymptotically free.

Lastly, for two sequences of random matrices  $(A_N)_N$  and  $(B_N)_N$  in  $(L^{\infty-}(\Omega) \otimes M_N(\mathbb{C}), \mathbb{E} \otimes \text{tr})_N$ , asymptotic freeness is equivalent to that any sequence of mixed cumulants, as defined via (2.20), vanish in the limit  $N \rightarrow \infty$ ; and almost sure asymptotic freeness is equivalent to that the mixed cumulants vanish almost surely in the limit  $N \rightarrow \infty$ .

Now we can now state Voiculescu’s theorem, that is central in bridging random matrix theory and free probability.

**Theorem 2** (Voiculescu [28]). *Let  $(A_N)_N$  and  $(B_N)_N$  be two sequences of  $N \times N$  random matrix ensembles in  $(L^{\infty-}(\Omega) \otimes M_N(\mathbb{C}), \mathbb{E} \otimes \text{tr})_N$  that admit asymptotic spectral distributions such that for any  $N$ , their matrix elements are independently distributed and  $A_N$  (or  $B_N$ ) is a unitarily invariant ensemble,<sup>80</sup> then  $(A_N)_N$  and  $(B_N)_N$  are almost surely asymptotically free.*

Let us revisit the matrix matrix model proposed in Section 3.3.2. We claimed that  $\tilde{R}_N$  and  $\tilde{R}$  model  $\tilde{r}_N$  and  $\tilde{r}$ , in the sense that the ESD of the former is well approximated by the spectral distribution of the later at large fixed values of  $N$  and  $k$ . We want to argue it by leveraging Voiculescu’s theorem 2. However, an obstacle is that we are not given a sequence of random matrices in (3.51) but rather a random matrix with fixed size. We do not have a sequence of random matrices so Voiculescu’s theorem doesn’t seem to be applicable here.

In order to make the contact with free probability, a standard solution is to “blow up” a finite matrix, say the deterministic  $k \times k$  radiation density matrix  $R$  given to us, by tensoring an identity factor of dimension  $L$ , therefore creating a sequence of  $kL \times kL$  matrices  $(R_{kL})_L := R \otimes I_L$  by adjusting  $L$ .<sup>81</sup> Note that the ESD of  $(R_{kL})_L$  remains the same and thus trivially converges to  $\nu_r$  as  $L \rightarrow \infty$ .

<sup>80</sup>A unitarily invariant ensemble means the joint distribution of the matrix elements is invariant under any unitary conjugation.

<sup>81</sup>More generally, in a (random) matrix polynomial involving standard independent random matrix ensembles, such as GUE, GE, CUE, and some deterministic matrices. One simply replaces the deterministic ones by some random variables  $\{d_i\}$  that share the same spectral distribution. Then the corresponding polynomial in  $\{d_i\}$  and the free semi-circular, circular or Haar elements is called the *free deterministic equivalent*, and it serves as a good approximation at large  $N$ . This procedure is often used to deal with (random) matrix polynomials involving single deterministic matrices or non-converging sequence of deterministic matrices. In our case, for example,  $r$  is a free deterministic equivalent of  $(R_{kL})_L$ . For a precise formulation of free deterministic equivalents, see Chapter 10.1 in Mingo-Speicher [21] and [64, 131, 132].

Consider now the matrix model (3.51). We can shift this factor  $N$  from  $C_N$  to  $R_N$  without changing the overall expression. Let's rewrite it as  $\sqrt{NR_N}C_N B_N C_N \sqrt{NR_N}$  where we have  $\mathbb{E}\text{tr}B_N = 1$ ,  $\text{tr}NR_N = 1$  and  $C_N$  has variance  $1/N$  instead of 1 as in (3.51). Then we blow up  $NR_N$  to sequences and  $R_{NL} := NR_N \otimes I_L$ . The Ginibre ensemble can be chosen to be of any shape so we can have the sequence of  $NL \times NL$  Ginibre matrices  $C_{NL}$ .

Since any unitary from left or right can be absorbed by the Ginibre ensemble, we can conjugate  $(B_{NL})_L$  by a Haar random unitary for free. Then it follows that  $(C_{NL})_L$ ,  $(B_{NL})_L$  and  $(R_{NL})_L$  are all independently distributed and any pair consists of one ensemble being unitarily invariant. Also, each ensemble admits a LSD. All the conditions of Voiculescu's theorem are thus satisfied and the almost sure asymptotic freeness follows.

As sequences of non-commutative random variables in  $(L^{\infty-}(\Omega) \otimes M_{NL}(\mathbb{C}), \mathbb{E} \otimes \text{tr})_L$ , we then have the convergence in the sense of Definition 6,  $(C_{NL})_L \rightarrow c$ ,  $(B_{NL})_L \rightarrow b_N$  and  $(R_{NL})_L \rightarrow r_N$  such that  $c, b_N, r_N$  are free from each other. This gives a concrete random matrix realization of the abstract non-commutative random variables we invoke in Section 3.3.1. In the next subsection, we discuss why the random matrix models we proposed capture the desired limiting spectral distributions.

### B.3 Sample covariance matrix and separable sample covariance matrix

It now makes sense to ask for the LSD for any polynomial of sequences a collection of asymptotically free random matrix ensembles. When each random matrix ensemble admits a LSD, the answer is given by the distribution of the corresponding non-commutative random variables at the large  $N$  limit. If the random matrices are almost surely asymptotically free, then the ESD of any polynomial almost surely converges weakly to the LSD. For the precise statement and its proof, see Lemma II.4.1 in [60].

For example, we are mostly interested in the product of two asymptotically free random matrices. Given two almost surely asymptotically free sequences of random positive<sup>82</sup> Hermitian matrices  $(A_N)_N$  and  $(B_N)_N$ , which have the LSD of  $\nu_a$  and  $\nu_b$  respectively, we almost surely have the sequence of ESDs  $(\nu_{\sqrt{A_N}B_N\sqrt{A_N}})_N$  converges weakly to the free multiplicative convolution  $\nu_a \boxtimes \nu_b$ ,

$$(\nu_{\sqrt{A_N}B_N\sqrt{A_N}})_N \xrightarrow{a.s.} \nu_a \boxtimes \nu_b . \quad (\text{B.10})$$

The same also holds for the free additive convolution.

We now look at the sample covariance matrix ensemble that models the free compound Poisson distribution, and the separable sample covariance matrix that models our free convolution formula (1.9).

The free Poisson distribution is modeled by the *Wishart ensemble*. Let  $X_N$  to be  $N \times M$  rectangular random matrices with i.i.d. complex Gaussian entries of zero mean and variance  $\alpha/M$ . Consider now the Wishart matrix

$$W_N = X_N X_N^* \in (L^{\infty-}(\Omega) \otimes M_N(\mathbb{C}), \mathbb{E} \otimes \text{tr}) . \quad (\text{B.11})$$

We demand that the ratio  $M/N \rightarrow \lambda$  as  $N, M \rightarrow \infty$ .

The resulting LSD is called the *Marchenko–Pastur (MP) distribution*, denoted as  $\mu_{\lambda, \alpha}^{\text{MP}}$ .

$$\mu_{\lambda, \alpha}^{\text{MP}} := \begin{cases} (1 - \lambda)\delta_0 + \mu, & \text{if } 0 \leq \lambda \leq 1 \\ \mu, & \text{if } \lambda > 1 \end{cases} \quad (\text{B.12})$$

where  $\mu(x) := \frac{\lambda}{2\pi\alpha x} \sqrt{(x_+ - x)(x - x_-)} \mathbf{1}_{[x_-, x_+]}$ ,  $x_{\pm} := \alpha(1 \pm 1/\sqrt{\lambda})^2$ .

<sup>82</sup>In fact, we only demand one of them, say  $A_N$ , to be positive.

Its moments reads

$$m_n(\mu_{\lambda,\alpha}^{\text{MP}}) = \lim_{N \rightarrow \infty} \text{tr} W_N^n = \alpha^n \sum_{k=1}^n \lambda^{-k} N(n, k) \quad (\text{B.13})$$

where  $N(n, k) := \binom{n}{k} \binom{n}{k-1} / n$  are the Narayana numbers. The free cumulants read

$$\kappa_n(\mu_{\lambda,\alpha}^{\text{MP}}) = \lambda^{1-n} \alpha^n . \quad (\text{B.14})$$

We can write the MP distribution in terms of a free Poisson distribution

$$\mu_{\lambda,\alpha}^{\text{MP}} = \mu_{\lambda,\alpha/\lambda} . \quad (\text{B.15})$$

We now show that a compound Poisson distribution is the LSD of the following random matrix ensemble,

$$X_N^* A_N X_N \in (L^{\infty-}(\Omega) \otimes M_N(\mathbb{C}), \mathbb{E} \otimes \text{tr}) , \quad (\text{B.16})$$

which consists of the same Ginibre ensemble as above and a sequence of independently distributed Hermitian matrices  $(A_N)_N$  that admit a LSD  $\nu$ .

It shares the same non-zero eigenvalues with a *sample covariance matrix*,  $\sqrt{A_N} X_N X_N^* \sqrt{A_N}$ . Their spectra are different when  $X_N$  is rectangular ( $N \times M$  with variance  $1/M$ ). Nonetheless, one can easily obtain one from the other via<sup>83</sup>

$$\nu_{X_N^* A_N X_N} = (N/M) \nu_{\sqrt{A_N} X_N X_N^* \sqrt{A_N}} + (1 - N/M) \delta_0 . \quad (\text{B.17})$$

If  $M/N \rightarrow \lambda$  as  $N, M \rightarrow \infty$  and  $\nu_{A_N} \rightarrow \nu$ , we claim that<sup>84</sup>

$$\begin{aligned} \nu_{X_N^* A_N X_N} &\xrightarrow{\text{a.s.}} \mu_{1/\lambda, \nu} , \\ \nu_{\sqrt{A_N} X_N X_N^* \sqrt{A_N}} &\xrightarrow{\text{a.s.}} \lambda \mu_{1/\lambda, \nu} + (1 - \lambda) \delta_0 . \end{aligned} \quad (\text{B.18})$$

We can use free probability to obtain this result. Since  $A_N$  is independent from  $X_N$ ,  $X_N$  is unitarily invariant and so is  $W_N = X_N X_N^*$  and both  $W_N$  and  $A_N$  admits the LSD's  $\mu_{\lambda,1}^{\text{MP}}$  and  $\nu$  respectively, they are almost surely asymptotically free as implied by Voiculescu's theorem 2. The LSD of  $(\sqrt{A_N} X_N X_N^* \sqrt{A_N})_N$  is given by the convolution  $\mu_{\lambda,1}^{\text{MP}} \boxtimes \nu$ .

Consider now the LSD of  $X_N^* A_N X_N$  and using (B.17) gives,

$$\nu_{X_N^* A_N X_N} \xrightarrow{\text{a.s.}} \lambda^{-1} \cdot \mu_{\lambda,\alpha}^{\text{MP}} \boxtimes \nu + (1 - \lambda^{-1}) \delta_0 . \quad (\text{B.19})$$

Let's evaluate the moments of the above LSD,

$$\begin{aligned} m_n(\lambda^{-1} \cdot \mu_{\lambda,1}^{\text{MP}} \boxtimes \nu + (1 - \lambda^{-1}) \delta_0) &= \lambda^{-1} \sum_{\pi \in \text{NC}_n} \left( \prod_{V \in \pi} \kappa_{|V|}(\mu_{\lambda,1}^{\text{MP}}) \cdot \prod_{\bar{V} \in \bar{\pi}} m_{|\bar{V}|}(\nu) \right) \\ &= \lambda^{-1} \sum_{\pi \in \text{NC}_n} \prod_{V \in \pi} \lambda^{1-|V|} \cdot \prod_{\bar{V} \in \bar{\pi}} m_{|\bar{V}|}(\nu) \\ &= \sum_{\pi \in \text{NC}_n} \prod_{V \in \pi} \lambda^{-1} m_{|V|}(\nu) = m_n(\mu_{1/\lambda, \nu}) , \end{aligned} \quad (\text{B.20})$$

<sup>83</sup>This formula holds in general for the ESDs  $\nu_{AB}$  and  $\nu_{BA}$  of two rectangular matrices  $A$  and  $B$  with sizes  $N \times M$  and  $M \times N$  respectively.

<sup>84</sup>This result was established by Silverstein and Bai for  $X_N$  being the more general Wigner ensemble [133].

where we used the formula (2.39) in the first line; in the second line we used the formula of free cumulants (B.14) with  $\alpha = 1$ ; and in the last step we used  $|\pi| + |\bar{\pi}| = 1 + n$ . Hence, the cumulants of the LSD is given by  $\kappa_n = \lambda^{-1} m_n(\nu)$ , which yields (B.18).

As for the sample covariance matrix (2.67), we can also consider the matrix ensemble that shares the same non-zero spectrum as (3.41), known as the *separable sample covariance matrix*

$$\sqrt{D_N} C_N^* A_N C_N \sqrt{D_N} \in (L^{\infty-}(\Omega) \otimes M_M(\mathbb{C}), \mathbb{E} \otimes \text{tr}) . \quad (\text{B.21})$$

where we have the a sequence of additional independently distributed random  $N \times N$  matrix  $(D_N)_N$  that admits a LSD  $\xi$ .  $C_N$  is an  $N \times N$  Ginibre matrix. We assume w.l.o.g. that  $M = N$  by allowing  $D_N$  to be rank-deficient.  $(D_N)_N$  is almost surely asymptotically free from the sample covariance matrix  $C_N^* A_N C_N$  according to Voiculescu's theorem. It follows from the above result on sample covariance matrix that

$$\nu_{\sqrt{D_N} C_N^* A_N C_N \sqrt{D_N}} \xrightarrow{a.s.} \xi \boxtimes \mu_{1,\nu} . \quad (\text{B.22})$$

In terms of the non-commutative random variables, we have the following convergence in distribution,

$$(\sqrt{D_N} C_N^* A_N C_N \sqrt{D_N})_N \longrightarrow \sqrt{d} c^* a c \sqrt{d} \quad (\text{B.23})$$

where  $(D_N)_N \rightarrow d$ ,  $(A_N)_N \rightarrow a$  and  $c$  is a circular element in some  $W^*$ -probability space  $(\mathcal{W}, \tau)$ .

Now we can verify that our matrix models in Section 3.3.2 correctly model the random variables in Section 3.3.1. We let  $D_N$  be the rescaled  $N \times N$  embedded radiation density matrix  $NR_N$ , and  $A_N$  be the black hole (random) thermal density matrix  $B_N$ . Then the ensemble (B.21) matches exactly with  $\tilde{R}_N$  in (3.51).

Based on the discussion in the end of last subsection, we fix  $N$  to be some large value and we use  $L$  for the varying parameter for the matrix size. We  $NR_N$  to a sequence, and define a sequence that extends (3.51) to the large  $L$  limit.

$$(\tilde{R}_{NL})_L := (N \sqrt{R_{NL}} C_{NL}^* B_{NL} C_{NL} \sqrt{R_{NL}})_L \in (L^{\infty-}(\Omega) \otimes M_{NL}(\mathbb{C}), \mathbb{E} \otimes \text{tr}) \quad (\text{B.24})$$

where at  $L = 1$  we have the  $N\tilde{R}_N$ . We need to put in the factor  $N$  to adjust the normalization.

Let the free deterministic equivalent of  $NR_N$  be  $r_N$  (cf. Footnote 81), in the sense that  $(R_{NL})_L \rightarrow r_N$ . Then (B.23) implies that  $\tilde{R}_N$  (3.51) models  $\tilde{r}_N$  (3.43), in the sense that  $(\tilde{R}_{NL})_L \rightarrow \tilde{r}_N$ . More explicitly, we have almost surely the sequence of ESD  $(\nu_{\tilde{R}_{NL}})_L$  converges weakly to  $\nu_{\tilde{r}_N}$ . Similarly, the arguments apply for  $\tilde{B}_N$  and  $\tilde{b}_N$ .

Lastly, we consider the large  $N$  limit. We have  $\tilde{R}_N \rightarrow \tilde{R}$  and  $p_N \tilde{r}_N p_N \rightarrow \tilde{r}$ . Because almost surely the sequence of ESD  $(\nu_{P_{NL} \tilde{R}_{NL} P_{NL}})_L$  converges weakly to  $\nu_{P_N \tilde{r}_N P_N}$ . If for any bounded continuous function  $f$  on  $\mathbb{R}$ , the convergence of the expectation value of  $f$  evaluated with the distribution  $(\nu_{P_{NL} \tilde{R}_{NL} P_{NL}})_L$  is uniform on  $N$ , we can then swap the limits to obtain  $\tilde{R} \rightarrow \tilde{r}$ .

## References

- [1] S. W. Hawking, *Particle creation by black holes*, *Communications in Mathematical Physics* **43** (1975) 199.
- [2] S. W. Hawking, *Breakdown of predictability in gravitational collapse*, *Physical Review D* **14** (1976) 2460.

- [3] G. Penington, *Entanglement wedge reconstruction and the information paradox*, *Journal of High Energy Physics* **2020** (2020) 1.
- [4] A. Almheiri, N. Engelhardt, D. Marolf and H. Maxfield, *The entropy of bulk quantum fields and the entanglement wedge of an evaporating black hole*, *Journal of High Energy Physics* **2019** (2019) 1.
- [5] A. Almheiri, R. Mahajan, J. Maldacena and Y. Zhao, *The Page curve of Hawking radiation from semiclassical geometry*, *Journal of High Energy Physics* **2020** (2020) 1.
- [6] G. Penington, S. H. Shenker, D. Stanford and Z. Yang, *Replica wormholes and the black hole interior*, *Journal of High Energy Physics* **2022** (2022) 1.
- [7] A. Almheiri, T. Hartman, J. Maldacena, E. Shaghoulian and A. Tajdini, *Replica wormholes and the entropy of Hawking radiation*, *Journal of High Energy Physics* **2020** (2020) 1.
- [8] A. Almheiri, T. Hartman, J. Maldacena, E. Shaghoulian and A. Tajdini, *The entropy of Hawking radiation*, *Rev. Mod. Phys.* **93** (2021) 035002.
- [9] D. Marolf and J. Polchinski, *Gauge-gravity duality and the black hole interior*, *Physical Review Letters* **111** (2013) 171301.
- [10] D. N. Page, *Average entropy of a subsystem*, *Physical Review Letters* **71** (1993) 1291.
- [11] D. N. Page, *Information in black hole radiation*, *Phys. Rev. Lett.* **71** (1993) 3743.
- [12] P. Calabrese and J. Cardy, *Entanglement entropy and quantum field theory*, *Journal of Statistical Mechanics: Theory and Experiment* **2004** (2004) P06002.
- [13] A. Lewkowycz and J. Maldacena, *Generalized gravitational entropy*, *Journal of High Energy Physics* **2013** (2013) 90.
- [14] T. Faulkner, A. Lewkowycz and J. Maldacena, *Quantum corrections to holographic entanglement entropy*, *Journal of High Energy Physics* **2013** (2013) 74.
- [15] N. Engelhardt and A. C. Wall, *Quantum extremal surfaces: holographic entanglement entropy beyond the classical regime*, *Journal of High Energy Physics* **2015** (2015) 73.
- [16] C. Akers and G. Penington, *Leading order corrections to the quantum extremal surface prescription*, *Journal of High Energy Physics* **2021** (2021) 1.
- [17] J. Wang, *The refined quantum extremal surface prescription from the asymptotic equipartition property*, *Quantum* **6** (2022) 655.
- [18] D. Harlow and D. Jafferis, *The factorization problem in Jackiw-Teitelboim gravity*, *Journal of High Energy Physics* **2020** (2020) 1.
- [19] R. Renner, *Security of quantum key distribution*, *International Journal of Quantum Information* **6** (2008) 1.
- [20] A. Nica and R. Speicher, *Lectures on the combinatorics of free probability*, vol. 13. Cambridge University Press, 2006.
- [21] J. A. Mingo and R. Speicher, *Free probability and random matrices*, vol. 35. Springer, 2017.
- [22] R. Speicher, *Lecture notes on Free Probability Theory*, *arXiv:1908.08125* (2019) .
- [23] J. Novak, *Three lectures on free probability*, *Random matrix theory, Interacting particle systems, and Integrable systems* **65** (2014) 13.
- [24] Z. Lixin, *Spectral analysis of large dimensional random matrices*, Ph.D. thesis, National University of Singapore, 2006.

- [25] W. Hachem, P. Loubaton and J. Najim, *The empirical distribution of the eigenvalues of a gram matrix with a given variance profile*, in *Annales de l'Institut Henri Poincaré (B) Probability and Statistics*, vol. 42, pp. 649–670, Elsevier, 2006.
- [26] D. Paul and J. W. Silverstein, *No eigenvalues outside the support of the limiting empirical spectral distribution of a separable covariance matrix*, *Journal of Multivariate Analysis* **100** (2009) 37.
- [27] R. Couillet and W. Hachem, *Analysis of the limiting spectral measure of large random matrices of the separable covariance type*, *Random Matrices: Theory and Applications* **3** (2014) 1450016.
- [28] D. Voiculescu, *Limit laws for random matrices and free products*, *Inventiones Mathematicae* **104** (1991) 201.
- [29] A. C. Wall, *Proof of the generalized second law for rapidly changing fields and arbitrary horizon slices*, *Physical Review D* **85** (2012) 104049.
- [30] P. W. Shor, *Equivalence of additivity questions in quantum information theory*, *Communications in Mathematical Physics* **246** (2004) 453.
- [31] M. B. Hastings, *Superadditivity of communication capacity using entangled inputs*, *Nature Physics* **5** (2009) 255.
- [32] M. Fukuda, T. Hasebe and S. Sato, *Additivity violation of quantum channels via strong convergence to semi-circular and circular elements*, *Random Matrices: Theory and Applications* **11** (2022) 2250012.
- [33] J. Wang, *On the information-theoretic aspects of black holes*, Ph.D. thesis, ETH Zurich, 2022.
- [34] J. Kudler-Flam, V. Narovlansky and S. Ryu, *Distinguishing random and black hole microstates*, *PRX Quantum* **2** (2021) 040340.
- [35] N. Cheng, C. Lancien, G. Penington, M. Walter and F. Witteveen, *Random tensor networks with nontrivial links*, *arXiv:2206.10482* (2022) .
- [36] T. Tao, *Topics in random matrix theory*, vol. 132. American Mathematical Soc., 2012.
- [37] V. F. Jones, *von Neumann algebras*, *Course at Berkeley University* (2003) .
- [38] D. Voiculescu, *Symmetries of some reduced free product  $C^*$ -algebras*, in *Operator algebras and their connections with topology and ergodic theory*, pp. 556–588, Springer, (1985).
- [39] D. Voiculescu, *Free probability and the von Neumann algebras of free groups*, *Reports on Mathematical Physics* **55** (2005) 127.
- [40] I. E. Segal, *A note on the concept of entropy*, *Journal of Mathematics and Mechanics* (1960) 623.
- [41] R. Longo and E. Witten, *A note on continuous entropy*, *arXiv preprint arXiv:2202.03357* (2022) .
- [42] E. Witten, *Aps medal for exceptional achievement in research: Invited article on entanglement properties of quantum field theory*, *Reviews of Modern Physics* **90** (2018) 045003.
- [43] I. Gelfand and M. Neumark, *On the imbedding of normed rings into the ring of operators in hilbert space*, *Matematicheskii Sbornik* **12** (1943) 197.
- [44] I. E. Segal, *Irreducible representations of operator algebras*, *Bulletin of the American Mathematical Society* **53** (1947) 73.
- [45] M. Rédei and S. J. Summers, *When are quantum systems operationally independent?*, *International Journal of Theoretical Physics* **49** (2010) 3250.
- [46] D. Voiculescu, *Addition of certain non-commuting random variables*, *Journal of functional analysis* **66** (1986) 323.

- [47] J. M. Lindsay and V. Pata, *Some weak laws of large numbers in noncommutative probability*, *Mathematische Zeitschrift* **226** (1997) 533.
- [48] U. Haagerup and S. Möller, *The law of large numbers for the free multiplicative convolution*, in *Operator algebra and dynamics*, pp. 157–186, Springer, (2013).
- [49] D. Voiculescu, *Multiplication of certain non-commuting random variables*, *Journal of Operator Theory* (1987) 223.
- [50] H. Bercovici and D.-V. Voiculescu, *Lévy-Hinčin type theorems for multiplicative and additive free convolution*, *Pacific journal of mathematics* **153** (1992) 217.
- [51] H. Bercovici, V. Pata and P. Biane, *Stable laws and domains of attraction in free probability theory*, *Annals of Mathematics* (1999) 1023.
- [52] G. Kreweras, *Sur les partitions non croisées d'un cycle*, *Discrete Mathematics* **1** (1972) 333.
- [53] P. Biane, *Some properties of crossings and partitions*, *Discrete Mathematics* **175** (1997) 41.
- [54] R. Speicher, *Multiplicative functions on the lattice of non-crossing partitions and free convolution*, *Mathematische Annalen* **298** (1994) 611.
- [55] H. Bercovici and D. Voiculescu, *Free convolution of measures with unbounded support*, *Indiana University Mathematics Journal* **42** (1993) 733.
- [56] N. R. Rao and R. Speicher, *Multiplication of free random variables and the  $s$ -transform: The case of vanishing mean*, *Electronic Communications in Probability* **12** (2007) 248.
- [57] J. Pennington, S. Schoenholz and S. Ganguli, *The emergence of spectral universality in deep networks*, in *International Conference on Artificial Intelligence and Statistics*, pp. 1924–1932, PMLR, 2018.
- [58] R. Chhaibi, T. Daouda and E. Kahn, *Free probability, newton lily pads and hyperbolicity of jacobians as a solution to the problem of tuning the architecture of neural networks*, *arXiv:2111.00841* (2021) .
- [59] S. T. Belinschi and H. Bercovici, *A new approach to subordination results in free probability*, *Journal d'Analyse Mathématique* **101** (2007) 357.
- [60] T. Mai, *On the analytic theory of non-commutative distributions in free probability*, Ph.D. thesis, Saarland University, 2017.
- [61] R. Speicher, *Combinatorial theory of the free product with amalgamation and operator-valued free probability theory*, vol. 627. American Mathematical Soc., 1998.
- [62] U. Haagerup and S. Thorbjørnsen, *A new application of random matrices:  $\text{Ext}(C_{\text{red}}^*(F_2))$  is not a group*, *Annals of Mathematics* (2005) 711.
- [63] J. W. Helton, T. Mai and R. Speicher, *Applications of realizations (aka linearizations) to free probability*, *Journal of Functional Analysis* **274** (2018) 1.
- [64] S. T. Belinschi, T. Mai and R. Speicher, *Analytic subordination theory of operator-valued free additive convolution and the solution of a general random matrix problem*, *Journal für die reine und angewandte Mathematik* **2017** (2017) 21.
- [65] S. T. Belinschi and H. Bercovici, *Partially defined semigroups relative to multiplicative free convolution*, *International Mathematics Research Notices* **2005** (2005) 65.
- [66] A. Nica and R. Speicher, *On the multiplication of free  $n$ -tuples of noncommutative random variables*, *American Journal of Mathematics* (1996) 799.
- [67] F. Mezzadri, *How to generate random matrices from the classical compact groups*, *Notices of the American Mathematical Society* **54** (2007) 592.

- [68] F. Hiai and D. Petz, *The semicircle law, free random variables and entropy*, *Mathematical Surveys & Monographs* (2006) .
- [69] Z. Yang, *The quantum gravity dynamics of near extremal black holes*, *Journal of High Energy Physics* **2019** (2019) 1.
- [70] P. Saad, S. H. Shenker and D. Stanford, *JT gravity as a matrix integral*, *arXiv:1903.11115* (2019) .
- [71] A. M. Tulino, S. Verdú et al., *Random matrix theory and wireless communications*, *Foundations and Trends in Communications and Information Theory* **1** (2004) 1.
- [72] R. Bousso and E. Wildenhain, *Gravity/Ensemble duality*, *Physical Review D* **102** (2020) 066005.
- [73] R. Bousso and M. Tomašević, *Unitarity from a smooth horizon?*, *Physical Review D* **102** (2020) 106019.
- [74] D. Marolf and H. Maxfield, *Transcending the ensemble: baby universes, spacetime wormholes, and the order and disorder of black hole information*, *Journal of High Energy Physics* **2020** (2020) 1.
- [75] D. Marolf and H. Maxfield, *Observations of hawking radiation: the page curve and baby universes*, *Journal of High Energy Physics* **2021** (2021) 1.
- [76] P. Hayden, S. Nezami, X.-L. Qi, N. Thomas, M. Walter and Z. Yang, *Holographic duality from random tensor networks*, *Journal of High Energy Physics* **2016** (2016) 1.
- [77] D. Harlow, *TASI lectures on the emergence of the bulk in AdS/CFT*, *arXiv:1802.01040* (2018) .
- [78] P. Hayden and G. Penington, *Learning the alpha-bits of black holes*, *Journal of High Energy Physics* **2019** (2019) 1.
- [79] C. Akers and G. Penington, *Quantum minimal surfaces from quantum error correction*, *arXiv:2109.14618* (2021) .
- [80] E. H. Lieb and M. B. Ruskai, *Proof of the strong subadditivity of quantum-mechanical entropy*, *Journal of Mathematical Physics* **14** (1973) 1938.
- [81] H. Araki, *Relative entropy of states of von neumann algebras*, *Publications of the Research Institute for Mathematical Sciences* **11** (1976) 809.
- [82] D. Petz, *Quasi-entropies for finite quantum systems*, *Reports on Mathematical Physics* **23** (1986) 57.
- [83] V. Chandrasekaran, R. Longo, G. Penington and E. Witten, *An algebra of observables for de sitter space*, *arXiv:2206.10780* (2022) .
- [84] J. D. Bekenstein, *Black holes and entropy*, *Phys. Rev. D* **7** (1973) 2333.
- [85] J. D. Bekenstein, *Universal upper bound on the entropy-to-energy ratio for bounded systems*, *Phys. Rev. D* **23** (1981) 287.
- [86] E. Witten, *Why does quantum field theory in curved spacetime make sense? and what happens to the algebra of observables in the thermodynamic limit?*, *arXiv:2112.11614* (2021) .
- [87] M. Tomamichel, *Quantum information processing with finite resources: mathematical foundations*, vol. 5. Springer, 2015.
- [88] R. Renner and S. Wolf, *Smooth Rényi entropy and applications*, in *International Symposium on Information Theory, 2004. ISIT 2004. Proceedings.*, p. 233, IEEE, 2004.
- [89] R. König, R. Renner and C. Schaffner, *The operational meaning of min-and max-entropy*, *IEEE Transactions on Information theory* **55** (2009) 4337.
- [90] M. Berta, *Single-shot quantum state merging*, *arXiv:0912.4495* (2009) .



- [91] N. Bao, G. Penington, J. Sorce and A. C. Wall, *Beyond toy models: distilling tensor networks in full AdS/CFT*, *Journal of High Energy Physics* **2019** (2019) 1.
- [92] C. Murthy and M. Srednicki, *Structure of chaotic eigenstates and their entanglement entropy*, *Physical Review E* **100** (2019) 022131.
- [93] D. Marolf, S. Wang and Z. Wang, *Probing phase transitions of holographic entanglement entropy with fixed area states*, *Journal of High Energy Physics* **2020** (2020) 1.
- [94] X. Dong and H. Wang, *Enhanced corrections near holographic entanglement transitions: a chaotic case study*, *Journal of High Energy Physics* **2020** (2020) 1.
- [95] P. Hayden and J. Preskill, *Black holes as mirrors: quantum information in random subsystems*, *Journal of High Energy Physics* **2007** (2007) 120.
- [96] A. Almheiri, X. Dong and D. Harlow, *Bulk locality and quantum error correction in AdS/CFT*, *Journal of High Energy Physics* **2015** (2015) 163.
- [97] X. Dong, D. Harlow and A. C. Wall, *Reconstruction of bulk operators within the entanglement wedge in gauge-gravity duality*, *Physical Review Letters* **117** (2016) 021601.
- [98] D. Harlow, *The ryu–takayanagi formula from quantum error correction*, *Communications in Mathematical Physics* **354** (2017) 865.
- [99] F. Pastawski, B. Yoshida, D. Harlow and J. Preskill, *Holographic quantum error-correcting codes: Toy models for the bulk/boundary correspondence*, *Journal of High Energy Physics* **2015** (2015) 1.
- [100] X. Dong, D. Harlow and D. Marolf, *Flat entanglement spectra in fixed-area states of quantum gravity*, *Journal of High Energy Physics* **2019** (2019) 1.
- [101] C. Akers and P. Rath, *Holographic Rényi entropy from quantum error correction*, *Journal of High Energy Physics* **2019** (2019) 1.
- [102] M. Berta, M. Christandl and R. Renner, *The quantum reverse shannon theorem based on one-shot information theory*, *Communications in Mathematical Physics* **306** (2011) 579.
- [103] F. Dupuis, M. Berta, J. Wullschleger and R. Renner, *One-shot decoupling*, *Communications in Mathematical Physics* **328** (2014) 251.
- [104] F. Dupuis, *The decoupling approach to quantum information theory*, *arXiv:1004.1641* (2010) .
- [105] A. Uhlmann, *The “transition probability” in the state space of a  $*$ -algebra*, *Reports on Mathematical Physics* **9** (1976) 273.
- [106] C. M. Caves, C. A. Fuchs and R. Schack, *Unknown quantum states: the quantum de Finetti representation*, *Journal of Mathematical Physics* **43** (2002) 4537.
- [107] R. König and R. Renner, *A de Finetti representation for finite symmetric quantum states*, *Journal of Mathematical physics* **46** (2005) 122108.
- [108] M. Christandl, R. König, G. Mitchison and R. Renner, *One-and-a-half quantum de Finetti theorems*, *Communications in Mathematical Physics* **273** (2007) 473.
- [109] R. Renner, *Symmetry of large physical systems implies independence of subsystems*, *Nature Physics* **3** (2007) 645.
- [110] R. Renner and J. Wang, *The black hole information puzzle and the quantum de Finetti theorem*, *arXiv:2110.14653* (2021) .
- [111] V. Chandrasekaran, G. Penington and E. Witten, *Large  $n$  algebras and generalized entropy*, *arXiv preprint arXiv:2209.10454* (2022) .

- [112] D. Stanford and E. Witten, *JT gravity and the ensembles of random matrix theory*, *arXiv:1907.03363* (2019) .
- [113] D. Stanford, *More quantum noise from wormholes*, *arXiv:2008.08570* (2020) .
- [114] C. V. Johnson, *Quantum gravity microstates from fredholm determinants*, *Physical Review Letters* **127** (2021) 181602.
- [115] R. Gopakumar and D. J. Gross, *Mastering the master field*, *Nuclear Physics B* **451** (1995) 379.
- [116] M. R. Douglas and M. Li, *Free variables and the two-matrix model*, *Physics Letters B* **348** (1995) 360.
- [117] M. R. Douglas, *Large  $n$  gauge theory—expansions and transitions*, *Nuclear Physics B Proceedings Supplements* **41** (1995) 66.
- [118] J. A. Mingo and R. Speicher, *Second order freeness and fluctuations of random matrices: I. gaussian and wishart matrices and cyclic fock spaces*, *Journal of Functional Analysis* **235** (2006) 226.
- [119] J. A. Mingo, P. Śniady and R. Speicher, *Second order freeness and fluctuations of random matrices: Ii. unitary random matrices*, *Advances in Mathematics* **209** (2007) 212.
- [120] B. Collins, J. A. Mingo, P. Śniady and R. Speicher, *Second order freeness and fluctuations of random matrices. iii: Higher order freeness and free cumulants.*, *Documenta Mathematica* **12** (2007) 1.
- [121] G. Borot, S. Charbonnier, E. Garcia-Failde, F. Leid and S. Shadrin, *Analytic theory of higher order free cumulants*, *arXiv:2112.12184* (2021) .
- [122] C. E. Shannon, *A mathematical theory of communication*, *The Bell system technical journal* **27** (1948) 379.
- [123] P. Hayden and A. Winter, *Counterexamples to the maximal  $p$ -norm multiplicativity conjecture for all  $p > 1$* , *Communications in Mathematical Physics* **284** (2008) 263.
- [124] F. G. Brandao and M. Horodecki, *On hastings' counterexamples to the minimum output entropy additivity conjecture*, *Open Systems & Information Dynamics* **17** (2010) 31.
- [125] M. Fukuda, *Revisiting additivity violation of quantum channels*, *Communications in Mathematical Physics* **332** (2014) 713.
- [126] S. T. Belinschi, B. Collins and I. Nechita, *Almost one bit violation for the additivity of the minimum output entropy*, *Communications in Mathematical Physics* **341** (2016) 885.
- [127] B. Collins, *Haagerup's inequality and additivity violation of the minimum output entropy*, *arXiv:1603.00577* (2016) .
- [128] B. Collins and I. Nechita, *Random matrix techniques in quantum information theory*, *Journal of Mathematical Physics* **57** (2016) 015215.
- [129] A. S. Holevo, *Complementary channels and the additivity problem*, *Theory of Probability & Its Applications* **51** (2007) 92.
- [130] P. Hayden and G. Penington, *Black hole microstates vs. the additivity conjectures*, *arXiv:2012.07861* (2020) .
- [131] R. Speicher and C. Vargas, *Free deterministic equivalents, rectangular random matrix models, and operator-valued free probability theory*, *Random Matrices: Theory and Applications* **1** (2012) 1150008.
- [132] C. Vargas Obieta, *Free probability theory: deterministic equivalents and combinatorics*, Ph.D. thesis, Saarland University, 2015.
- [133] J. W. Silverstein and Z. Bai, *On the empirical distribution of eigenvalues of a class of large dimensional random matrices*, *Journal of Multivariate analysis* **54** (1995) 175.