entropy

MDPI

Article

A Max-Flow Approach to Random Tensor Networks

Khurshed Fitter [1], Faedi Loulidi [2] and Ion Nechita [3,*]

[1] Quantum Science and Engineering Department, Ecole Polytechnique Federal de Lausane, 1015 Lausanne, Switzerland; khurshed.fitter@epfl.ch
[2] Networked Quantum Devices Unit, Okinawa Institute of Science and Technology Graduate University, Onna-son 904-0495, Okinawa, Japan; faedi.loulidi@oist.jp
[3] Laboratoire de Physique Théorique, Université de Toulouse, CNRS, UPS, 31062 Toulouse, France
* Correspondence: nechita@irsamc.ups-tlse.fr

Abstract

The entanglement entropy of a random tensor network (RTN) is studied using tools from free probability theory. Random tensor networks are simple toy models that help in understanding the entanglement behavior of a boundary region in the anti-de Sitter/conformal field theory (AdS/CFT) context. These can be regarded as specific probabilistic models for tensors with particular geometry dictated by a graph (or network) structure. First, we introduce a model of RTN obtained by contracting maximally entangled states (corresponding to the edges of the graph) on the tensor product of Gaussian tensors (corresponding to the vertices of the graph). The entanglement spectrum of the resulting random state is analyzed along a given bipartition of the local Hilbert spaces. The limiting eigenvalue distribution of the reduced density operator of the RTN state is provided in the limit of large local dimension. This limiting value is described through a maximum flow optimization problem in a new graph corresponding to the geometry of the RTN and the given bipartition. In the case of series-parallel graphs, an explicit formula for the limiting eigenvalue distribution is provided using classical and free multiplicative convolutions. The physical implications of these results are discussed, allowing the analysis to move beyond the semiclassical regime without any cut assumption, specifically in terms of finite corrections to the average entanglement entropy of the RTN.

Keywords: random tensor networks (RTN); entanglement entropy; max-flow min-cut theorem; free probability theory; AdS/CFT correspondence; Ryu–Takayanagi formula; series-parallel graphs; free convolution; Marchenko–Pastur distribution

check for updates

Received: 26 May 2025
Revised: 26 June 2025
Accepted: 27 June 2025
Published: 15 July 2025

Citation: Fitter, K.; Loulidi, F.; Nechita, I. A Max-Flow Approach to Random Tensor Networks. *Entropy* **2025**, *27*, 756. https://doi.org/10.3390/e27070756

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The AdS/CFT correspondence is described as a quantum theory (more precisely, a *conformal field theory*) lying on the boundary of an *anti-de Sitter* spacetime geometry [1]. Many particular features of this correspondence remain mysterious, in particular the link with quantum information theory and *entanglement*. It was shown in [2] that for a fixed time slice, the entanglement behavior of a given region of the boundary quantum theory is proportional to the minimal hypersurface bulk area homologous to the region of interest (known as the Ryu–Takayanagi entanglement entropy). In the context of AdS/CFT, the Ryu-Takayanagi formula demonstrates a crucial link between the entanglement behavior of an intrinsic quantum theory and its connection with the bulk gravitational field. These results open up a new perspective on the understanding of quantum gravity in the AdS/CFT

framework from the viewpoint of entanglement and quantum information theory. For a complete introduction, see [3] and references therein.

The difficulty of computing the entanglement properties of boundary quantum theories has led to the development of attractable simple models, particularly the *tensor network* and *random tensor network* frameworks. Initially, the tensor network framework started with "good" models approximating ground states in condensed matter physics. In the context of condensed matter physics, tensor networks represent ground states of a class of gapped Hamiltonians [4]. Moreover, tensor networks have paved the way to understanding different physical properties such as the classification of topological phases of matter. An extensive review of all the different applications can be found in [4]. Recently, other extensions of tensor networks to random tensor networks for studying random matrix product states or projected entangled pairs of states have been introduced in [5–7]. The random tensor network (or simply RTN) was initiated in [8] as a toy model reproducing the key properties of entanglement behavior in the AdS/CFT context [9–15]. Moreover, the random tensor network framework has appeared in different active areas of research ranging from condensed matter physics to random quantum circuits and measurement frameworks [16–24].

In general, a random tensor network (or simply RTN) is defined as a random quantum state via a given fixed graph structure, as will be described below. The main problem is to compute the *average entanglement* between two complementary regions of the graph as $D \to \infty$, where D denotes the dimension of the Hilbert space of the model. Various results have been established, allowing the entanglement entropy of RTN models to be understood as toy models that mimic the entanglement behavior in quantum gravity. The scaling of the entanglement entropy (in the $D \to \infty$ limit) as a function of the size and *number of minimal cuts* needed to separate the region of interest from the rest of the graph has been explored in different works [8,25]. Moreover, several directions have been explored that move beyond the toy model picture of (random) tensor networks [26,27].

The focus of this work is a *maximal flow* approach to the analysis of general random tensor networks. The use of the maximal flow approach was already explored in [10] to compute the entanglement negativity, in [28] to derive the Ryu–Takayanagi entanglement entropy in the continuum setting, and in Appendix C of [8] in a moment computation similar to the ones discussed in this work; see also [9,25]. As described in the previous paragraph, the model consists of defining a random quantum state from a given fixed graph structure. This model consider a graph with edges (bulk edges) and half-edges (boundary edges). The use of bulk and boundary edges will become clear from the definition of the model. A finite-dimensional Hilbert space $\mathbb{C}^D$ is associated with each half-edge and a tensor product Hilbert space $(\mathbb{C}^D)^{\otimes 2}$ with each edge. The edge Hilbert spaces generate a local Hilbert space associated with each vertex of the graph. In order to define an RTN, each component of the graph is associated with a quantum state generated at random. For this, a random Gaussian state is generated for each vertex and a *maximally entangled state* is associated with each of the edges. A *random tensor network* is defined by projecting all of the maximally entangled states associated with the graph onto the total random states generated in each vertex. The obtained random tensor network lies in the full-boundary Hilbert space. The main goal of this work is to consider a sub-boundary region A of the graph and evaluate the entanglement behavior of the associated residual state as $D \to \infty$. The first computation is to estimate the moment computation of the state associated with the region A as $D \to \infty$. With the help of the *maximal flow*, which is developed in detail in this work, the moments are estimated *without any assumption regarding cuts*, and it is shown that they converge to the moment of a graph-dependent measure. It is shown that if the obtained *partial order* is *series-parallel*, the measure associated with the graph

can be explicitly constructed *without any cut assumption* through the use of *free probability theory*. Moreover, the *existence of higher-order correction terms* of the entanglement entropy (provided by a graph-dependent measure) is demonstrated, and these can be explicitly provided if the partial order is series-parallel. In different examples, it is shown how the measures associated with the initial graph can be computed explicitly in the case where the obtained partial order is series-parallel. The link between quantum information theory, free probability, and random tensor networks was already explored in [25]. There, a general link state was used to represent the effect of the bulk matter field in the AdS/CFT context, extending the semiclassical regime with correction terms of the entanglement structure. Moreover, the obtained results in [25] *assumed* the existence of two disjoint minimal cuts separating region A from the rest. In this work, only maximally entangled states in the bulk edge of the model are used; higher-order correction terms are obtained without any cut assumption, which may be interpreted as *intrinsic fluctuations*. In the context of AdS/CFT, these are intrinsic to the quantum spacetime nature of bulk gravitational field without any bulk matter field.

The main differences between this work and the previous literature [8,25] are highlighted below. Through the use of the maximum flow approach, we emphasize the importance of the partial order defined by disjoint paths achieving the maximum flow to the computation of the correction term for the entropy of entanglement of the random tensor network states. Indeed, both the minimum cut and the maximum flow perspectives provide the dominating term ($\sim \log D$). However, performing the analysis with the maximum flow allows for a combinatorial characterization of the moments of the reduced state. Moreover, in the case where the flow partial order is *series-parallel*, an explicit formula for the correction term is provided in terms of a probability measure constructed from the network structure using both the free and classical multiplicative convolutions from (free) probability theory.

The rest of this work is organized as follows. Section 2 provides a summary of the main results. Section 3 introduces the random tensor network framework. Section 4 presents the moment computation of a given state ρ_A associated with a given sub-boundary region of the graph A. Section 5 uses the maximal flow approach to compute the asymptotic scaling of the moments and demonstrates convergence to a graph-dependent measure. Section 6 introduces the notion of series-parallel partial order and uses the free probability to explicitly show how a graph-dependent measure can be constructed with the free product convolution and classical measurement product. Section 7 provides various examples of random tensor networks and explicitly shows the associated measure obtained in the case of an obtained partial order that is series-parallel. Section 8 presents the main technical results, which use concentration inequalities to show that the obtained higher-order entanglement correction terms are graph-dependent without any cut assumption; moreover, the graph-dependent measure can be explicitly constructed if the partial order is series-parallel.

2. Main Results

This section introduces the main definitions and results obtained in this work. The focus is placed on the computation of the entanglement entropy of a given random tensor network. The most general framework of random tensor networks is considered, and the entanglement structure of the random tensor is studied with respect to a fixed bipartition of the total Hilbert space. By addressing the problem through a *network flow approach*, the leading term can be computed along with the higher-order correction terms of the entanglement entropy, which are *graph-dependent*. The higher-order correction terms are found to play a crucial role in various areas, particularly in the context of AdS/CFT [8,25,27,29],

as will be discussed following the presentation of the main results. The key results of this work can be informally summarized as follows.

> In the limit of large local Hilbert space dimension D, the average Rényi entanglement entropy of an RTN G across a given bipartition $(A|B)$ has the following:
>
> - A dominating term of the form $\text{maxflow}(G_{A|B}) \cdot \log D$
> - A finite correction term which is graph-dependent
>
> In the case where $G_{A|B}$ is a *series-parallel* graph, the distribution of the entanglement spectrum (and consequently the finite entropy correction) can be computed as an iterative classical and free convolution of Marčhenko–Pastur distributions.

A random tensor network is associated with a corresponding random quantum state $|\psi_G\rangle$ that encodes the structure of a graph G. To describe this, the graph G and relevant terminology are introduced below. Further technical details and definitions of the model are provided in Section 3. Let $G = (V, E)$ denote a connected undirected finite graph with (full) edges and half-edges (sometimes referred to as an *open graph*); the former encode the internal entanglement structure of the quantum state $|\psi_G\rangle$, while the latter represent the physical systems (Hilbert spaces) on which $|\psi_G\rangle$ is supported. The sets of edges (bulk edges) and half-edges (boundary edges) are denoted by E_b and E_∂, respectively. Formally, the sets of edges and half-edges are provided by $E_b := \{e_{x,y} \,|\, e_{x,y} = (x, y) : x, y \in V\}$ and $E_\partial := \{e_x = (x, \cdot) : x \in V\}$, where $E := E_b \sqcup E_\partial$. The corresponding *random tensor* $|\psi_G\rangle$ is then defined as

$$|\psi_G\rangle := \Big\langle \bigotimes_{e \in E_b} \Omega_e \,\Big|\, \bigotimes_{x \in V} g_x \Big\rangle \in (\mathbb{C}^D)^{\otimes |E_\partial|},$$

where $|g_x\rangle$ are taken to be random Gaussian states defined in the local Hilbert space of each vertex x. Moreover, a maximally entangled state $|\Omega_e\rangle \in \mathbb{C}^D \otimes \mathbb{C}^D$ is associated with each (full) edge $e \in E_b$, which serves to contract the internal degrees of freedom of the tensor network. A representation of a random tensor network is provided in Figure 1, which is discussed in detail to illustrate the main results of this work. Further details are provided in Definition 3.

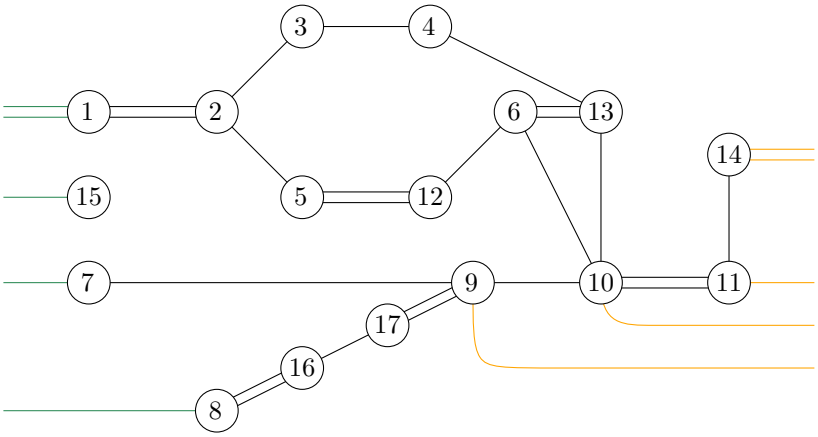


Figure 1. A tensor network depicting a tensor from $(\mathbb{C}^D)^{\otimes 10}$ obtained by contracting 17 tensors. The ten Hilbert space factors are partitioned into two subsets $B \sqcup A$.

As mentioned earlier, the objective of this work is to evaluate the *entanglement entropy* of the random quantum state $|\psi_G\rangle$ along a bipartition $A|B$ of the boundary edges $E_\partial = A \sqcup B$. This analysis is carried out in the limit of the large local Hilbert space dimension $D \to \infty$. To evaluate the entanglement entropy of the pure state $|\psi_G\rangle$, the asymptotic *entanglement*

spectrum along the bipartition $A|B$ is computed, that is, the limiting spectrum of the density matrix $\rho_A = \mathrm{Tr}_B|\psi_G\rangle\langle\psi_G|$. From this spectral information, the average Rényi entanglement and von Neumann entropies for the approximate normalized state $\tilde{\rho}_A$ can be deduced, respectively provided by:

$$\tilde{\rho}_A := D^{-|E_\partial|}\rho_A \to \begin{cases} \lim_{D\to\infty} \mathbb{E}\, S_n(\tilde{\rho}_A) & \text{with} \quad S_n(\rho) := \frac{1}{1-n}\log(\mathrm{Tr}\,\rho^n), \\ \lim_{D\to\infty} \mathbb{E}\, S(\tilde{\rho}_A) & \text{with} \quad S(\rho) := -\mathrm{Tr}(\rho\log\rho). \end{cases}$$

Above, the expectation is taken with respect to the Gaussian distribution of the independent random tensors $|g_x\rangle$ present at each vertex of the graph. The use of approximate normalized state instead of a correctly normalized state $\rho_A/\,\mathrm{Tr}\,\rho_A$ will become clear from Section 8.

First, the *moments* of the random matrix ρ_A are computed exactly, then the main contributing terms at large dimensions are analyzed by relating the problem to a *maximum flow* question in a related graph. By employing the maximal flow and tools from *free probability theory*, the leading and fluctuating terms of the Rényi entropy are derived and the behavior of the von Neumann entanglement entropy is subsequently deduced.

Moment computation. First, the un-normalized state ρ_A is considered and its moments are computed. As a first step, the *graphical Wick formula* from [30] is used to obtain

$$\mathbb{E}[\mathrm{Tr}(\rho_A^n)] = \sum_{\alpha=(\alpha_x)\in\mathcal{S}_n^{|V|}} D^{n|E_\partial|-H_G^{(n)}(\alpha)}, \ \forall\, n\in\mathbb{N}, \tag{1}$$

where $H_G^{(n)}(\alpha)$ can be understood as the Hamiltonian of a classical "spin system" in which each spin variable takes a value from the permutation group $\mathcal{S}_n$:

$$H_G^{(n)}(\alpha) := \sum_{(x,\cdot)\in A} |\gamma_x^{-1}\alpha_x| + \sum_{(x,\cdot)\in B} |\,\mathrm{id}_x^{-1}\,\alpha_x| + \sum_{(x,y)\in E_b} |\alpha_x^{-1}\alpha_y|.$$

In the above, the identity permutation $\mathrm{id}_x \in \mathcal{S}_n$ is associated with the region B (corresponding to taking the partial trace over B), while the full-cycle permutation $\gamma_x = (n\ n-1\ \cdots\ 2\ 1)$ is associated with the region A (corresponding to the trace of the n-th power of ρ_A). A more precise statement and proof can be found in Proposition 1 in Section 4. It should also be noted that the contribution of the normalization term of $\tilde{\rho}_A$ is provided by

$$\mathbb{E}[(\mathrm{Tr}\,\rho_A)^n] = \sum_{\alpha=(\alpha_x)\in\mathcal{S}_n^{|V|}} D^{n|E_\partial|-h_G^{(n)}(\alpha)}, \ \forall\, n\in\mathbb{N}$$

where

$$h_G^{(n)}(\alpha) := \sum_{(x,\cdot)\in E_\partial} |\,\mathrm{id}_x^{-1}\,\alpha_x| + \sum_{(x,y)\in E_b} |\alpha_x^{-1}\alpha_y|.$$

Note that in the above formula the quantity $h_G^{(n)}(\alpha)$ is simply $H_G^{(n)}(\alpha)$ with $A = \emptyset$; see Proposition 2 for more details. Note that in the particular case with $n = 2$, the authors of [8] provided an exact mapping to the partition function of a classical Ising model. Notice the frustrated boundary conditions of the Hamiltonian above; vertices connected to region A prefer the configuration $\alpha_x = \gamma_x$, while vertices connected to region B prefer the low energy state $\alpha_x = \mathrm{id}_x$. The connection between the combinatorics of the dominating terms in the random matrix model considered here and an effective spin system Hamiltonian have already been extensively studied in the quantum statistical mechanics literature; see, e.g., [22,31,32].

Maximal flow. The *(max)-flow* approach consists of identifying the leading terms from the moment formula above as $D \to \infty$. For this, a *network* $G_{A|B}$ is introduced. This network is derived from the original graph G by connecting all of the half-edges in A to an extra vertex γ (sink) and all of the half-edges in B to id (source). In $G_{A|B}$, the vertices are valued in the permutation group $\mathcal{S}_n$, with all the half-edges connected either to the source id or to the sink γ. The flow approach consists of looking at the different paths starting from the source id to the sink γ. The different paths in the flow approach induce an ordering structure (more precisely a *poset structure*) in the network $G_{A|B}$. Intuitively, the maximal flow will consist of searching the maximal number of these paths such that the source and sink will be disconnected upon removing the paths from the network. More precisely, by *Menger's theorem*, the maximum flow in this graph is equal to the number of edge-disjoint *augmenting paths* that start from the source id and end in the sink γ. Figure 7 represents the different paths achieving maximal flow in the network $G_{A|B}$ from the original graph G as represented in Figure 1. This procedure allows us to find a *lower bound* to the Hamiltonian $H^{(n)}_{G_{A|B}}(\alpha)$ that can be attained by some choice of the variables α_x.

Theorem 1. *For all $n \geq 1$, one has*

$$\min_{\alpha \in \mathcal{S}_n^{|V|}} H^{(n)}_{G_{A|B}}(\alpha) = (n-1)\,\mathrm{maxflow}(G_{A|B}),$$

where $H^{(n)}_{G_{A|B}}(\alpha)$ denotes the extended Hamiltonian in the network $G_{A|B}$. After all augmenting paths achieving the maximum flow in $G_{A|B}$ have been removed, a clustered graph $G^c_{A|B}$ *remains, which is formed by clustering all the remaining connected components (see Figure 8). Importantly, the maximality of the flow implies that the cluster-vertices* $[\mathrm{id}]$ *and* $[\gamma]$ *in this clustered graph are disjoint. Further details and the proof of the above result are provided in Proposition 5.*

As a direct consequence of the above result, the moment convergence as $D \to \infty$ can be deduced; further details of the following result are provided in Theorem 4.

Theorem 2. *In the limit $D \to \infty$, for all $n \geq 1$ one has*

$$\lim_{D\to\infty} \mathbb{E}\frac{1}{D^{F(G_{A|B})}}\left[\mathrm{Tr}\left(\left(D^{F(G_{A|B})-|E_\partial|}\,\rho_A\right)^n\right)\right] = m_n,$$

where m_n are the moments of a probability measure $\mu_{G_{A|B}}$ and $F(G_{A|B}) = \mathrm{maxflow}(G_{A|B})$.

Moreover, we can show that the normalization term converges to 1, as shown in Corollary 1. The previous maximum flow computation provides the first order in the formula for the average entanglement entropy of random tensor network states:

$$\mathbb{E}[S_n(\tilde{\rho}_A)] \approx \mathrm{maxflow}(G_{A|B}) \cdot \log D \;\forall\; n \geq 1.$$

Free probability theory and entanglement. The main contribution of this work is to show that the *second order* (or the finite corrections) of the Rényi and von Neumann entanglement entropy can be found by carefully analyzing the set of augmenting paths achieving the maximum flow in the graph $G_{A|B}$. When the different paths achieve the maximal flow in the graph $G_{A|B}$, a *partial order* $G^o_{A|B}$ is obtained after the clustering operation, where the vertices are the different permutation clusters formed from the clustered graph $G^c_{A|B}$. Figure 9 shows the obtained partial order from the original graph G in Figure 1. The results are general, and become explicit in the setting where the partial order $G^o_{A|B}$ is *series-parallel*.

With the help of *free probability theory*, the second-order correction terms of each of the Rényi and von Neumann entropy can be deduced in this setting.

Definition 1. *A graph G is called* series-parallel *if it can be constructed recursively using the following two operations:*

- *Series concatenation: $G = H_1 \,|\mathbb{S}|\, H_2$ is obtained by identifying the sink of H_1 with the source of H_2.*
- *Parallel concatenation: $G = H_1 \,|\mathbb{P}|\, H_2$ obtained by identifying the sources and sinks of H_1 and H_2.*

Definition 2. *A probability measure μ_G is associated with a series-parallel graph G, defined recursively as follows:*

- *The Dirac mass at 1 is associated with the trivial graph $G_{triv} = (\{s,t\},\{\{s,t\}\})$, as follows: $\mu_{G_{triv}} := \delta_1$.*
- *Series concatenation: $\mu_{G\,|\mathbb{S}|\,H} := \mu_G \boxtimes \mathrm{MP} \boxtimes \mu_H$. Here*

$$\mathrm{dMP} := \frac{1}{2\pi}\sqrt{4t^{-1}-1}\,\mathrm{d}t$$

  is the Marčhenko–Pastur distribution and $\boxtimes$ is the free convolution product. For more details, see Appendix A.
- *Parallel concatenation: $\mu_{G\,|\mathbb{P}|\,H} := \mu_G \times \mu_H$.*

Theorem 3. *In the limit $D \to \infty$, the average Rényi entanglement entropy $\forall\ n \geq 1$ and von Neumann entropy of an approximate normalized state $\tilde{\rho}_A := D^{-|E_\partial|}\rho_A$ respectively behave as follows:*

$$\mathbb{E}[S_n(\tilde{\rho}_A)] = \mathrm{maxflow}(G_{A|B}) \cdot \log D - \frac{1}{n-1}\log \int t^n \,\mathrm{d}\mu_{G_{A|B}}(t) + o(1),$$
$$\mathbb{E}[S(\tilde{\rho}_A)] = \mathrm{maxflow}(G_{A|B}) \cdot \log D - \int t\,\log t\,\mathrm{d}\mu_{G_{A|B}} + o(1).$$

For further details and the proof of the above statements, the reader is referred to Corollary 2. In particular, if the obtained partial order $G^o_{A|B}$ is series-parallel, then the measure $\mu_{G_{A|B}} = \mu_{G^o_{A|B}}$ can be explicitly constructed, as detailed in Theorem 5. The use of the approximate normalized state rather than the normalized state $\tilde{\rho}_A := \rho_A / \operatorname{Tr}\rho_A$ is justified by the concentration result for $\operatorname{Tr}\rho_A$ presented in Section 8.1.

It was previously argued in [8,27] that if one wants to encode the *quantum fluctuations*, then instead of a maximally entangled state it is necessary to use a general "link state" $|\varphi_e\rangle$ defined by

$$e \in E_b \to |\varphi_e\rangle := \sum_{i=1}^{D} \sqrt{\lambda_{e,i}}\,|i_x, i_y\rangle.$$

It was recently shown in [25] that the quantum fluctuations beyond the semiclassical regime in AdS/CFT are obtained for the non-flat spectra of the link state under the existence *assumption* of *two minimal cuts*. The use of a generic link state in the context of AdS/CFT represents the bulk matter field contribution.In this work, the *maximal flow* approach has been used to demonstrate the existence of quantum fluctuations without any *minimal cut assumption* and with the *maximally entangled state* serving as the link state. The higher-order correction terms obtained in this context can be interpreted as the "intrinsic" quantum fluctuations of spacetime geometry in the absence of any bulk matter field represented by a general link state.

For example, in the case of the graph represented in Figure 1, the resulting partial order $G^o_{A|B}$ is series-parallel (see Figure 9), where

$$G^o_{A|B} = G_1 \boxed{\mathsf{S}} G_2 \boxed{\mathsf{S}} G_3 \quad \text{with} \quad \mu_{G^o_{A|B}} = \mu_{G_1} \boxtimes \mathrm{MP} \boxtimes \mu_{G_2} \boxtimes \mathrm{MP} \boxtimes \mu_{G_3} = \mu_{G_1} \boxtimes \mathrm{MP}^{\boxtimes 2}.$$

As represented in Figure 10, the graphs G_2 and G_3 are trivial; hence, $\mu_{G_2} = \mu_{G_3} = \delta_1$. The graph G_1 can be factored as a parallel composition of two other graphs, as represented in Figure 11:

$$G_1 = G_5 \boxed{\mathsf{P}} G_4 \quad \text{with} \quad \mu_{G_1} = \mu_{G_4} \times \mu_{G_5}.$$

The graph G_4, as represented in Figure 12, factorizes as follows:

$$G_4 = \left(G_6 \boxed{\mathsf{P}} G_7\right) \boxed{\mathsf{S}} G_8 \quad \text{with} \quad \mu_{G_4} = \left(\mu_{G_6} \times \mu_{G_7}\right) \boxtimes \mathrm{MP} \boxtimes \mu_{G_8} = (\mathrm{MP} \times \mathrm{MP}) \boxtimes \mathrm{MP}$$

where we use the fact that G_6 and G_7 are series compositions of two trivial graphs; thus, $\mu_{G_6} = \mu_{G_7} = \mathrm{MP}$, while $\mu_{G_8} = \delta_1$.

Moreover, the graph G_5, as represented in Figure 13, factorizes as follows:

$$G_5 = G_9 \boxed{\mathsf{S}} G_{10} \boxed{\mathsf{S}} \left(G_{11} \boxed{\mathsf{P}} G_{12}\right) \boxed{\mathsf{S}} G_{13}$$

with the associated measure

$$\mu_{G_5} = \mu_{G_9} \boxtimes \mathrm{MP} \boxtimes \mu_{G_{10}} \boxtimes \mathrm{MP} \boxtimes \left(\mu_{G_{11}} \times \mu_{G_{12}}\right) \boxtimes \mathrm{MP} \boxtimes \mu_{G_{13}} = \mathrm{MP}^{\boxtimes 3} \boxtimes \left(\mathrm{MP}^{\boxtimes 2} \times \mathrm{MP}\right),$$

where the series composition for G_{11} and G_{12} has been used iteratively, with their respective measures provided by $\mu_{G_{11}} = \mathrm{MP}^{\boxtimes 2}$ and $\mu_{G_{12}} = \mathrm{MP}$. In the case of random tensor network represented in Figure 1, the partial order is series-parallel, with the associated measure

$$G^o_{A|B} = G_1 \boxed{\mathsf{S}} G_2 \boxed{\mathsf{S}} G_3,$$

with

$$\mu_{G^o_{A|B}} = \left\{\left[\mathrm{MP}^{\boxtimes 3} \boxtimes \left(\mathrm{MP}^{\boxtimes 2} \times \mathrm{MP}\right)\right] \times \left[(\mathrm{MP} \times \mathrm{MP}) \boxtimes \mathrm{MP}\right]\right\} \boxtimes \mathrm{MP}^{\boxtimes 2},$$

which is obtained by combining all the results stated above. The minimal cuts associated with the network $G_{A|B}$ (see Figure 7) can also be considered, as represented in Figure 14. It should be noted that several minimal cuts exist; for simplicity, four ways of achieving the minimal cut in the network are represented. Note that some of the cuts share common edges.

3. Random Tensor Networks

In this section, starting from a given (open) graph with edges (bulk edges) and half-edges (boundary edges), the random tensor network model is introduced. For this purpose, a Hilbert space is associated with each edge and half-edge of the graph. The edge Hilbert spaces induce a local Hilbert space for each vertex in the graph. A random Gaussian state is assigned to each of the vertices, and a maximally entangled state is associated with each edge. The random tensor network is defined by projecting all of the maximally entangled states associated with the edges of the graph onto the vertex states, which are provided by the tensor product of all the random Gaussian vectors. The main definitions of the model are introduced in this section, and the different entanglement notions are recalled as well.

Section 3.1 introduces the random tensor network model, after which Section 3.2 recalls the different entanglement notions and their properties.

3.1. Random Tensor Network

The construction of the random tensor network model is presented in the following. Let $G = (V, E)$ be a bulk connected undirected finite open graph with edges and half-edges. The multi-sets of edges and half-edges are denoted by E_b and E_∂, respectively; multiple edges are allowed in the graph. Formally, bulk edges $e \in E_b$ are of the form $e = (x, y)$ for some $x, y \in V$, while half-edges $e' \in E_\partial$ are of the form $e' = (x, \cdot)$ for some $x \in V$, where the $\cdot$ indicates that e' is a partial edge. The total set of edges E is decomposed as

$$E := E_b \sqcup E_\partial.$$

Bulk connectivity is assumed, meaning that all vertices in the bulk region of the graph are connected; this corresponds to the "connected network" property as defined in Definition 2 of [33]. The cardinalities of the bulk, boundary, and total edge multi-sets are denoted by $|E_b|$, $|E_\partial|$, and $|E| = |E_b| + |E_\partial|$, respectively.

A Hilbert space $\mathbb{C}^D$ is associated with each half-edge on a given vertex in the graph, while the Hilbert space $\mathbb{C}^D \otimes \mathbb{C}^D$ is assigned to each bulk edge connecting two vertices, where D is a finite parameter known as the *bond dimension*. A random Gaussian vector is defined for each vertex of the graph state, lying in the local Hilbert space associated with that vertex. Furthermore, a maximally entangled state is associated with each edge of the graph. Thus, the *random tensor network* is defined as a random quantum state constructed by projecting the total tensor product of the random Gaussian states for each vertex onto all of the maximally entangled states formed along the bulk edges (see Definition 3).

Formally, Hilbert spaces are associated with each part of the graph G as follows:

- A finite-dimensional Hilbert space $\mathcal{H}_{e_x}$ is assigned to each half-edge defined on a vertex x:

$$\begin{aligned} e_x \in E_\partial \to \mathcal{H}_{e_x} &:= \mathbb{C}^D \\ E_\partial \to \mathcal{H}_\partial &:= \bigotimes_{e_x \in E_\partial} \mathcal{H}_{e_x}. \end{aligned}$$

- A Hilbert space $\mathcal{H}_{e_{x,y}}$ is associated with each edge $e_{x,y} \in E_b$:

$$e_{x,y} \in E_b \to \mathcal{H}_{e_{x,y}} = \mathcal{H}_{e\to x} \otimes \mathcal{H}_{e\to y} := \mathbb{C}^D \otimes \mathbb{C}^D,$$

  where $\mathcal{H}_{e_{x,y}}$ denotes the Hilbert space connecting the two vertices x and y, which is composed as a tensor product of the two corresponding half-edges.
- For each vertex $x \in V$, the local vertex Hilbert space $\mathcal{H}_x$ is defined as follows:

$$\begin{aligned} x \in V \to \mathcal{H}_x &:= \bigotimes_{E \ni e \to x} \mathcal{H}_{e\to x} \\ V \to \mathcal{H}_V &:= \bigotimes_{x\in V} \mathcal{H}_x = \bigotimes_{x\in V} \bigotimes_{E\ni e\to x} \mathcal{H}_{e\to x}. \end{aligned}$$

  Above, the Hilbert space $\mathcal{H}_x$ represents the local Hilbert space associated with a vertex x defined as all the edges of Hilbert space that contribute locally; the notation $e \to x$ denotes that the edge e is incident to the vertex x.

Having defined the general Hilbert space structure associated with a generic graph G, in the following we define quantum states in the graph G in order to introduce the

random tensor network model. Each vertex $x \in V$ is associated with a random vector (un-normalized quantum state) $|g_x\rangle \in \mathcal{H}_x$ sampled from an *i.i.d. Gaussian distribution*:

$$\begin{aligned} x \in V &\to |g_x\rangle \in \mathcal{H}_x \\ V &\to \bigotimes_{x \in V} |g_x\rangle \in \mathcal{H}_V. \end{aligned}$$

A normalized maximally entangled state $|\Omega_e\rangle$ is also associated with each bulk edge $e_{x,y} \in E_b$, provided by

$$\begin{aligned} e_{x,y} \in E_b &\to \mathcal{H}_{e_{x,y}} \\ e_{x,y} &\to |\Omega_e\rangle := \frac{1}{\sqrt{D}} \sum_{i=1}^{D} |i_x, i_y\rangle, \end{aligned} \tag{2}$$

where $|i_x\rangle$ and $|i_y\rangle$ indicate the state associated with the vertex x sharing an edge with y.

Definition 3. *A* random tensor network $|\psi_G\rangle$ *is defined as a projection of the vertex state over all of the maximally entangled states* $|\Omega_e\rangle$ *for each* $e_{x,y}$ *in* E_b*, where*

$$|\psi_G\rangle := \left\langle \bigotimes_{e \in E_b} \Omega_e \,\middle|\, \bigotimes_{x \in V} g_x \right\rangle \in \left(\mathbb{C}^D\right)^{\otimes |E_\partial|}. \tag{3}$$

It should be mentioned that the following example is used in all other parts of this work as an illustration of the different results obtained in each section.

Example 1. *An illustration of a random tensor network is provided in Figure 1, where the boundary region of G is provided by all the half-edges* $E_\partial := \{e_1, e_{15}, e_7, e_8, e_9, e_{10}, e_{11}, e_{14}\}$*. In Figure 1, the region A is defined as the set of half-edges at the vertices* $\{9, 10, 11, 14\}$*, that is,* $A := \{e_9, e_{10}, e_{11}, e_{14}\}$*. The complementary region* $B := E_\partial \setminus A$ *consists of the half-edges associated with the vertices* $\{1, 15, 7, 8\}$*, namely,* $B := \{e_1, e_{15}, e_7, e_8\}$*.*

It should also be mentioned that in the present construction of the random tensor network, the edges and half-edges are used to generate the vertex Hilbert space $\mathcal{H}_x$. Other types of random tensor network models have already been explored in the literature; see [8,9,25] and the references therein. In the previously mentioned models, the bulk and boundary vertices are first defined, whereas in the present approach the focus is placed on the edges and half-edges, which generate the local Hilbert space for each vertex, with the bulk states provided by a maximally entangled state. The initial work on random tensor networks was presented in [8], where the aim was to compute the entanglement entropy of a subregion of the random tensor network; as the bond dimension tends to infinity, this becomes proportional to the minimal cuts of the graph, reproducing the famous Ryu–Takayanagi entanglement entropy [2] in a discrete version.

In a recent work [25], the authors associated a state with a general "link" state connecting two bulk vertices, thereby generalizing previous models in which the existence of two non-crossing minimal cuts was allowed. This result allowed them to compute higher-order correction terms of the entanglement entropy. In this work, our main goal is to derive the higher-order correction terms using the maximal flow approach without any minimal cut assumption, with maximally entangled states connecting the bulk vertices.

3.2. Entanglement

In the following, different entanglement notions used in quantum information theory are recalled, in particular the von Neumann entropy and Rényi entropy.

The von Neumann entropy for a given normalized quantum state ρ is defined as

$$S(\rho) := -\operatorname{Tr}(\rho \log \rho). \tag{4}$$

In physical systems with an exponential number of degrees of freedom, the computation is generally difficult. A generalization exists in which diagonalization of the density matrix ρ is not required. This definition, introduced by Rényi, is known as the Rényi entropy and defined as

$$S_n(\rho) := \frac{1}{1-n} \log(\operatorname{Tr} \rho^n), \tag{5}$$

where it is well known that the Rényi entropy converges to the von Neumann entropy as $n \to 1$. The definitions provided above are for normalized quantum states; if the state is not normalized, one should first normalize it and then compute the entropy.

A subtlety regarding the upper bound on the rank of the reduced density matrix induced by the minimal cut should be mentioned here. A minimal cut is defined as the minimal number of edges in a graph that must be removed in order to fully separate a given fixed region of the graph. Although it is straightforward to observe that the rank of the reduced density matrix ρ_A is upper-bounded by the local dimension D raised to the number of edges in the set A, i.e., $\operatorname{rank}(\rho_A) \leq D^{|A|}$, a subtlety arises in that the rank of the reduced density matrix is upper-bounded by the minimum number of connecting edges, or the bottleneck (min-cut), rather than by the number of edges:

$$\operatorname{rank}(\rho_A) \leq D^{F_A} \tag{6}$$

where F_A is the min-cut (or the number of edges in the "bottleneck"); see, e.g., [8].

This can be demonstrated more clearly using an example. Consider a state $|\psi_G\rangle$ from which ρ_A can be constructed, as shown in Figure 2.

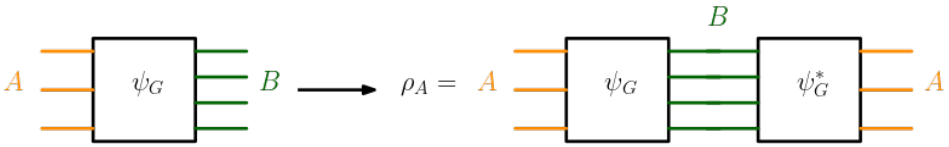


Figure 2. Constructing ρ_A from $|\psi_G\rangle$ by tracing out the edges corresponding to the set B.

The internal structure of $|\psi_G\rangle$ can be considered by dividing the graph into two subgraphs, denoted as L and R, which are connected by the "bottleneck", that is, the set of all edges for which their removal would disconnect the boundary sets A and B, see Figure 3.

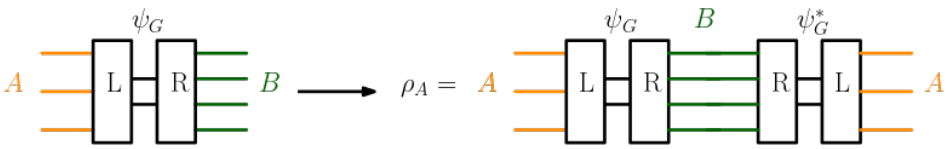


Figure 3. A factorization of the tensor ψ_G yields a upper bound on the rank of the corresponding reduced density matrix ρ_A.

Now, it is clear that $\operatorname{rank}(\rho_A) \leq D^{F_A}$, where in this case $F_A = 2$; consequently,

$$S(\rho_A) \leq F_A \log D. \tag{7}$$

Having established the natural intuition around the role of the min-cut (F_A) in upper-bounding the entropy, the (maximal) flow approach for the random tensor network will be developed in the following sections.

4. Moment Computation

Given a random tensor network, we next investigate the behavior of entanglement between a specified subregion of the network and the remainder. To this end, we first address the moment computation of the quantum state $\tilde{\rho}_A$ for a subregion $A \subseteq E_\partial$. This initial computation will subsequently enable our analysis of the Rényi and von Neumann entropies in the following sections.

Let $A \subseteq E_\partial$ be a sub-boundary region of the graph G. The complementary region of A is denoted by $B := E_\partial \setminus A$. The Hilbert spaces associated with the boundary regions A and B are denoted by $\mathcal{H}_A := \bigotimes_{e_x \in A} \mathcal{H}_{e_x}$ and $\mathcal{H}_B := \bigotimes_{e_x \in B} \mathcal{H}_{e_x}$, respectively.

In this work, we computed the *average entanglement entropy* at large bond dimensions:

$$\tilde{\rho}_A := \frac{\rho_A}{\operatorname{Tr} \rho_A} \to \begin{cases} \lim_{D\to\infty} \mathbb{E}\, S_n(\tilde{\rho}_A) \\ \lim_{D\to\infty} \mathbb{E}\, S(\tilde{\rho}_A) \end{cases} \tag{8}$$

where $\tilde{\rho}_A$ is the normalized quantum state obtained by tracing out the region B, i.e., $\rho_A = \mathrm{Tr}_B |\psi_G\rangle\langle\psi_G|$, with the partial trace over the Hilbert space $\mathcal{H}_B$. In the above expression, the average is taken over all of the random Gaussian tensors at the vertices.

The first computation to be addressed here is the moment computation, as described in the following proposition. This computation will subsequently enable our calculation of the average entanglement entropy (Rényi and von Neumann entropies) in the limit $D \to \infty$, as analyzed in detail in the following sections. The result has previously been obtained in a very similar setting by Hastings [33] (Theorem 3, E_{ind} ensemble); see also Section 5 of [8] for a closely related derivation.

Proposition 1. *For any $A \subseteq E_\partial$, it holds that*

$$\mathbb{E}[\mathrm{Tr}(\rho_A^n)] = \sum_{\alpha=(\alpha_x)\in \mathcal{S}_n^{|V|}} D^{n|E|-n|E_b|-H_G^{(n)}(\alpha)}, \ \forall\, n \in \mathbb{N}, \tag{9}$$

where $H_G^{(n)}(\alpha)$ can be understood as the Hamiltonian of a classical "spin system" in which each spin variable takes a value from the permutation group $\mathcal{S}_n$:

$$H_G^{(n)}(\alpha) := \sum_{(x,\cdot)\in A} |\gamma_x^{-1}\alpha_x| + \sum_{(x,\cdot)\in B} |\,\mathrm{id}_x^{-1}\,\alpha_x| + \sum_{(x,y)\in E_b} |\alpha_x^{-1}\alpha_y|. \tag{10}$$

Before providing the proof of the above proposition, we recall some properties of the permutation group $\mathcal{S}_n$ and fixed the relevant notation. Here, the symbol γ_x is used to denote the *total cycle* in the permutation group $\mathcal{S}_n$ evaluated at $(x,\cdot) \in A$:

$$\forall (x,\cdot) \in A, \qquad \gamma_x = (n \ldots 1).$$

Recall that a notion of distance in $\mathcal{S}_n$, known as the *Cayley distance*, can be defined by

$$\begin{aligned} &\mathcal{S}_n \times \mathcal{S}_n \to \mathbb{R}^+ \\ &d : (\alpha_i, \alpha_j) \to d(\alpha_i, \alpha_j) := n - \#(\alpha_i^{-1}\alpha_j), \end{aligned}$$

where $\#(\alpha)$ stands for the number of cycles in α. The distance $d(\alpha_i, \alpha_j)$ provides the minimum number of transpositions required to turn α_i into α_j. In general, the distance in $\mathcal{S}_n$ satisfies the triangle inequality, where

$$d(\alpha_i, \alpha_j) \leq d(\alpha_i, \sigma) + d(\sigma, \alpha_j).$$

In particular, σ is said to be a *geodesic* between α_i and α_j in $\mathcal{S}_n$ if $d(\alpha_i, \alpha_j) = d(\alpha_i, \sigma) + d(\sigma, \alpha_j)$. The following notation for the distance is adopted instead of $d(\cdot, \cdot)$, where

$$(\alpha_i, \alpha_j) \in \mathcal{S}_n \times \mathcal{S}_n, \quad d(\alpha_i, \alpha_j) = |\alpha_i^{-1}\alpha_j|.$$

Proof. To prove the result announced in the proposition, it should first be remarked that we can use the well-known replica trick to write the trace on the left-hand side of Equation (9) as follows:

$$\mathrm{Tr}(\rho_A^n) = \mathrm{Tr}\left(|\psi_G\rangle\langle\psi_G|^{\otimes n} U_{\gamma_A} \otimes \mathrm{id}_B\right).$$

The trace on the left-hand side (on $\mathcal{H}_A$) may be rewritten as a full trace on n copies of the complete (bulk and boundary) Hilbert space, as shown on the right-hand side of the equation above. The notation $U_{\gamma_A} = \bigotimes_{(x,\cdot)\in A} U_{\gamma_x}$ is used to denote the tensor product of unitary representations of the permutation $\gamma_x = (n \dots 1) \in \mathcal{S}_n$ for each half-edge $(x, \cdot) \in A$.

By expanding and taking the average over random Gaussian tensors, we obtain

$$\begin{aligned}
\mathbb{E}\,\mathrm{Tr}(\rho_A^n) &= \mathrm{Tr}\left(\mathbb{E}\Big[|\psi_G\rangle\langle\psi_G|^{\otimes n}\Big]\, U_{\gamma_A} \otimes \mathrm{id}_B\right) \\
&= \mathrm{Tr}\left(\bigotimes_{e\in E_b} |\Omega_e\rangle\langle\Omega_e|^{\otimes n} \mathbb{E}\Big[\bigotimes_{x\in V} |g_x\rangle\langle g_x|^{\otimes n}\Big]\, U_{\gamma_A} \otimes \mathrm{id}_B\right) \\
&= \mathrm{Tr}\left(\bigotimes_{e\in E_b} |\Omega_e\rangle\langle\Omega_e|^{\otimes n} \bigotimes_{x\in V} \mathbb{E}\Big[|g_x\rangle\langle g_x|^{\otimes n}\Big]\, U_{\gamma_A}\right),
\end{aligned}$$

where the shorthand notation U_{γ_A} in the last equation above is used in place of $U_{\gamma_A} \otimes \mathrm{id}_B$. We recall the following property of random Gaussian states (see [34]):

$$\forall x \in V, \quad \mathbb{E}\Big[|g_x\rangle\langle g_x|^{\otimes n}\Big] = \sum_{\{\alpha_x\}\in\mathcal{S}_n} U_{\alpha_x}$$

where U_{α_x} is the unitary representation of $\alpha_x \in \mathcal{S}_n$. Each permutation $\alpha_x \in \mathcal{S}_n$ acts on each vertex in Hilbert space; hence, each acts implicitly on each half-edge incident to the vertex $x \in V$. Thus, the moment formula becomes

$$\begin{aligned}
\mathbb{E}\,\mathrm{Tr}(\rho_A^n) &= \sum_{\{\alpha_x\}\in\mathcal{S}_n} \mathrm{Tr}\Big[\bigotimes_{e\in E_b} |\Omega_e\rangle\langle\Omega_e|^{\otimes n} \bigotimes_{x\in V} U_{\alpha_x}\, U_{\gamma_A}\Big] \\
&= D^{-n|E_b|} \sum_{\{\alpha_x\}\in\mathcal{S}_n} \prod_{(x,\cdot)\in A} D^{\#(\gamma_x^{-1}\alpha_x)} \prod_{(x,\cdot)\in B} D^{\#(\mathrm{id}_x^{-1}\alpha_x)} \prod_{(x,y)\in E_b} D^{\#(\alpha_x^{-1}\alpha_y)},
\end{aligned}$$

where the number of loops obtained by contracting the maximally entangled states (edges) upon taking the trace is counted by the formula above. The factor $D^{-n|E_b|}$ arises from the normalization associated with the contraction of the bulk edges (see Equation (2), where each edge contributes a factor of D^{-1}). By employing the relation between the Cayley distance and the number of loops, the result stated in the proposition is obtained. □

Graphically, the formula can be interpreted using Figure 4, where the case with $n = 3$ is illustrated. By employing the graphical integration technique for Wick integrals as presented in [30,35], loops are obtained. Consequently, Cayley distances of three types arise: (a) between id_x and elements directly connected to it, originating from the region B; (b) between γ_x and elements directly connected to it, originating from the region A; and (c) between elements neither directly connected to id nor γ, originating from the bulk. Based on this, the Hamiltonian can be rewritten in terms of Cayley distances, as follows:

$$H_G^{(n)}(\alpha) := \sum_{(x,\cdot)\in A} |\gamma_x^{-1}\alpha_x| + \sum_{(x,\cdot)\in B} |\,\mathrm{id}_x^{-1}\,\alpha_x| + \sum_{(x,y)\in E_b} |\alpha_x^{-1}\alpha_y| \tag{11}$$

where $(x, \cdot) \in A$ represents the half-edges in A, $(x, \cdot) \in B$, represents the half-edges in B, and $(x, y) \in E_b$ represents the edges in the bulk of the tensor network.

In the above proposition, only the numerator term of the normalized quantum state $\tilde{\rho}_A$ have been addressed. However, to compute the von Neumann and Rényi entropies (see Equations (4) and (5)), the state must be normalized and the moment computed.

The following proposition provides the moment computation of the normalization term in $\tilde{\rho}_A$.

Proposition 2. *For any $A \subseteq E_\partial$, it is found that*

$$\mathbb{E}\left[(\operatorname{Tr}\rho_A)^n\right] = \sum_{\alpha=(\alpha_x)\in\mathcal{S}_n^{|V|}} D^{n|E|-n|E_b|-h_G^{(n)}(\alpha)}, \ \forall\, n \in \mathbb{N}, \tag{12}$$

where the Hamiltonian $h_G^{(n)}(\alpha)$ is provided by

$$h_G^{(n)}(\alpha) := \sum_{(x,\cdot)\in E_\partial} |\operatorname{id}_x^{-1}\alpha_x| + \sum_{(x,y)\in E_b} |\alpha_x^{-1}\alpha_y|. \tag{13}$$

Proof. The proof of this proposition follows directly from Proposition 1 by taking $A = \emptyset$ such that $h_G^{(n)}(\alpha)$ is obtained as the special case of $H_G^{(n)}(\alpha)$ with $A = \emptyset$. □

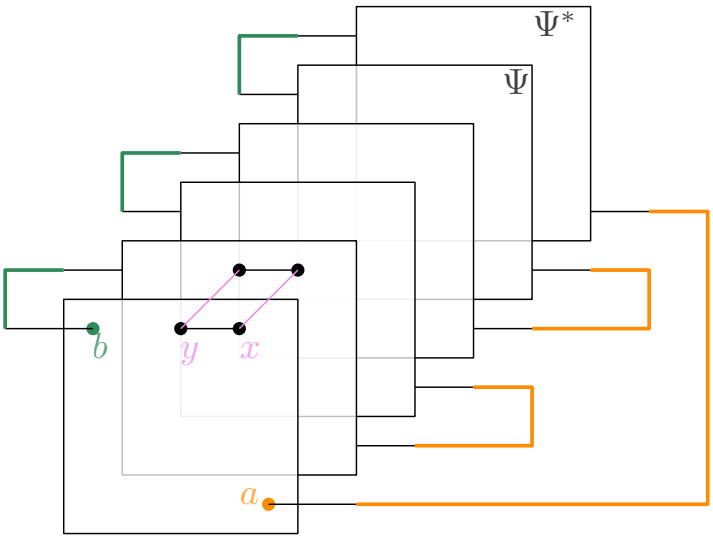


Figure 4. Graphical representation of the Wick theorem for the moment computation with $n = 3$.

5. Asymptotic Behavior of Moments

In this section, the leading contributing terms of the moment as $D \to \infty$ are described using the *(maximal) flow* approach. First, the (maximal) flow approach is introduced, which allows for the estimation of the leading terms of the moments as $D \to \infty$; further details can be found in Proposition 5. Based on this result, the convergence of the moment as $D \to \infty$ to the moments of a graph-dependent measure $\mu_{G_{A|B}}$ can be deduced, as detailed in Theorem 4.

First, the results from the previous section are recalled. In Proposition 1, it was shown that the moments are provided by

$$\mathbb{E}\operatorname{Tr}(\rho_A^n) = \sum_{\alpha=(\alpha_x)\in\mathcal{S}_n^{|V|}} D^{n|E|-n|E_b|-H_G^{(n)}(\alpha)},$$

where the spin-valued Hamiltonian in the permutation group $\mathcal{S}_n$ is provided by

$$H_G^{(n)}(\alpha) := \sum_{(x,\cdot)\in A} |\gamma_x^{-1}\alpha_x| + \sum_{(x,\cdot)\in B} |\,\mathrm{id}_x^{-1}\,\alpha_x| + \sum_{(x,y)\in E_b} |\alpha_x^{-1}\alpha_y|.$$

In particular, the contribution of the normalization term in $\tilde{\rho}_A$ (see Equation (8)) is the extended Hamiltonian $h_G^{(n)}(\alpha)$, as shown in Proposition 2 when one takes $A = \emptyset$ in $H_G^{(n)}(\alpha)$.

The main goal of this section is to analyze the main contributing terms of the moment as $D \to \infty$. The leading terms will consist of solving the following minimization problem:

$$\min_{\alpha\in\mathcal{S}_n^{|V|}} H_G^{(n)}(\alpha).$$

In particular, $h_G^{(n)}(\alpha)$ is minimized, yielding the leading contributing term as $D \to \infty$ for the normalization term of $\tilde{\rho}_A$. The minimization problem described above enables deduction of the moment convergence as $D \to \infty$ to the moment of the graph-dependent measure $\mu_{G_{A|B}}$ in Theorem 4.

The minimization problem above is addressed using the *(maximal)-flow* approach. This approach consists of first constructing a *network* $G_{A|B}$ from the original graph G. This network is constructed by first adding two extra vertices γ and id to G in such a way that all the half-edges associated with A are connected to the total cycle γ and that half-edges in B are connected to id. The network $G_{A|B}$ has the same bulk structure of G, with the difference that all the vertices in $G_{A|B}$ are valued in the permutation group $\mathcal{S}_n$.

The flow approach is based on the identification of different augmenting paths in the network $G_{A|B}$, each starting from id and ending at γ. These various paths induce an *order* structure in $G_{A|B}$. By removing all augmenting paths in $G_{A|B}$, a lower bound for the extended Hamiltonian $H_{G_{A|B}}^{(n)}(\alpha)$ can be established; further details are provided in Proposition 3. Furthermore, it is demonstrated that the minimum is attained when the maximal flow from id to γ is achieved, as discussed in Proposition 5. In particular, it is shown that the minimum of the extended Hamiltonian $h_{G_{A|B}}^{(n)}(\alpha)$ is zero; see Proposition 6 for additional details.

Before the flow approach is introduced, it should be noted that the contributing terms of the moments at large dimensions were previously analyzed in [25] using the *(minimal) cut* approach. In that work, the existence of two *disjoint minimal cuts* in the graph (separating the region of interest from the remainder of the graph) was assumed to determine the contributions at large bond dimensions. In contrast, in the maximal flow approach presented here, no assumption regarding the existence of (minimal) cuts is made. By identifying various augmenting paths that achieve the maximal flow and applying the well-known maximal-flow minimal-cut theorem (see, e.g., [36], Theorem 8.6), the different minimal cuts can be deduced without any prior assumption.

Definition 4. *Let the network* $G_{A|B} = (\tilde{V}, \tilde{E})$ *be defined from the initial graph* $G = (V, E)$ *such that*

$$\tilde{V} := V \sqcup \{\mathrm{id}, \gamma\} \quad \textit{and} \quad \tilde{E} := E_{\tilde{A}} \sqcup E_b \sqcup E_{\tilde{B}},$$

where the regions $E_{\tilde{A}}$ *and* $E_{\tilde{B}}$ *is defined as follows:*

$$E_{\tilde{A}} := \bigsqcup_{x\in V_A} (x, \gamma)$$
$$E_{\tilde{B}} := \bigsqcup_{x\in V_B} (\mathrm{id}, x)$$

where V_A and V_B respectively denote all of the vertices associated with the boundary region A and B. Moreover the vertices are valued in the permutation group $\mathcal{S}_n$, where

$$\forall x \in \tilde{V} \to \alpha_x \in \mathcal{S}_n.$$

It should be remarked that in the above definition, the graph $G_{A|B}$ is constructed in such a way that all of the half-edges $(x, \cdot) \in A$ are connected to $\gamma = (n \cdots 1) \in \mathcal{S}_n$ and that the half-edges $(x, \cdot) \in B$ are connected to id. Note also that there are no half-edges in $G_{A|B}$; the bulk region in the network $G_{A|B}$ remains the same as the one in graph G.

The extended Hamiltonian $H^{(n)}_{G_{A|B}}(\alpha)$ of $H^{(n)}_{G}(\alpha)$ in the network $G_{A|B}$ is provided by

$$H^{(n)}_{G_{A|B}}(\alpha) := \sum_{x \in V_A} |\gamma^{-1}\alpha_x| + \sum_{x \in V_B} |\,\mathrm{id}^{-1}\,\alpha_x| + \sum_{(x,y) \in V_b} |\alpha_x^{-1}\alpha_y|, \tag{14}$$

where each term in the new Hamiltonian is valued in the network $G_{A|B}$. Moreover, the sums in the above formula are over the vertices V_A, V_B, and V_b, which are the vertices with the respective half-edges in the region A, B, and E_b.

As mentioned earlier, the flow approach consists of analyzing different paths that start from id and end in γ. This induces a natural orientation of the network $G_{A|B}$ (more precisely, a *poset structure*). In the following, the sets of different (edge-disjoint) paths in $G_{A|B}$ is defined.

Definition 5. *Let $\mathcal{P}(G_{A|B})$ be the set of all possible paths from the source to the sink in $G_{A|B}$, where the source and sink are* id *and γ, respectively. Formally, the set of paths $\mathcal{P}(G_{A|B})$ is defined as*

$$\mathcal{P}(G_{A|B}) := \{\pi_i : \quad \pi_i : \mathrm{id} \to \gamma\},$$

where $\{\pi_i\}_i$ are all the paths connecting id *to γ.*

Definition 6. *Let $\tilde{\mathcal{P}}(G_{A|B})$ be the set of all families consisting of* edge-disjoint paths *in $\mathcal{P}(G_{A|B})$:*

$$\tilde{\mathcal{P}}(G_{A|B}) := \{\pi_i \in \mathcal{P}(G_{A|B}) : \{\pi_i\}_i \text{ are edge-disjoint }\}.$$

Remark 1. *It is clear from this definition that $\tilde{\mathcal{P}}(G_{A|B}) \subseteq \mathcal{P}(G_{A|B})$.*

The process of searching for different paths that start from id and end at γ induces an ordering, more precisely a *poset structure*, in the network $G_{A|B}$. In the following, a definition of a poset structure is provided which will later be utilized in the maximal flow approach to minimize $H^{(n)}_{G_{A|B}}(\alpha)$.

Definition 7. *The poset structure $\mathcal{P}_o(G_{A|B})$ is a* homogeneous *relation denoted by $\leq$ and satisfying the following conditions for all $\alpha_x, \alpha_y, \alpha_z \in \tilde{V}$:*

- *Reflexivity: $\alpha_x \leq \alpha_x$;*
- *Antisymmetry: $\alpha_x \leq \alpha_y$ and $\alpha_y \leq \alpha_x$, implying $\alpha_x = \alpha_y$;*
- *Transitivity: $\alpha_x \leq \alpha_y$ and $\alpha_y \leq \alpha_z$, implying $\alpha_x \leq \alpha_z$.*

Definition 8. *The* natural ordering *is defined as*

$$\mathrm{id} \leq \alpha_1 \leq \alpha_2 \leq \cdots \leq \alpha_n \leq \gamma$$

for a path $\pi_i \in \mathcal{P}(G_{A|B})$ *provided by*

$$\pi_i : \mathrm{id} \to \alpha_1 \to \alpha_2 \to \cdots \to \alpha_n \to \gamma.$$

Another useful notion in (maximal) flow analysis is the *permutation cluster*. A permutation cluster of a given permutation α_x is defined as all of the edge-connected permutations to α_x.

Definition 9. *A* permutation cluster $[\alpha_x]$ *is defined as all the edge-connected permutations to* $\alpha_x \in \mathcal{S}_n$.

Remark 2. *With the poset structure in* $G_{A|B}$*, a naturally-induced ordering is obtained in the cluster structures for each permutation connected to the permutation elements* $\{\alpha_i\}_{i\in\{x,y,z\}}$*, where all the properties of the above definition can be extended to the cluster* $[\alpha_i]$ *of a given permutation* α_i*. More precisely, the following hold:*

$$\begin{aligned} \alpha_x \leq \alpha_y &\implies [\alpha_x] \leq [\alpha_y]; \\ \alpha_x = \alpha_y &\implies [\alpha_x] = [\alpha_y]; \\ \alpha_x \leq \alpha_y \leq \alpha_z &\implies [\alpha_x] \leq [\alpha_y] \leq [\alpha_z]. \end{aligned}$$

Definition 10. *The max flow in* $G_{A|B}$ *is the maximum of all of the edge-disjoint paths in* $\tilde{\mathcal{P}}(G_{A|B})$*:*

$$\mathrm{maxflow}(G_{A|B}) := \max\Big\{|\pi| \ : \ \pi \in \tilde{\mathcal{P}}(G_{A|B}) \text{ is a set of edge-disjoint paths}\Big\}.$$

The following proposition provides a lower bound of the extended Hamiltonian $H_{G_{A|B}}(\alpha)$ that is saturated when the maximal flow in $G_{A|B}$ is achieved, as shown in Proposition 5. Similar ideas have been used by Hastings in Lemma 4 of [33] to lower-bound the moments of a random tensor network map.

Proposition 3. *Let* $\alpha \in \mathcal{S}_n^{|V|}$ *and* $\tilde{\mathcal{P}}(G_{A|B})$ *be an arbitrary set of edge-disjoint paths in* $G_{A|B}$ *and set* $k := |\tilde{\mathcal{P}}(G_{A|B})|$*; then, the following inequality holds:*

$$H^{(n)}_{G_{A|B}}(\alpha) \geq k(n-1) + H^{(n)}_{G_{A|B}\setminus\bigsqcup_{i\in[k]}\pi_i}(\alpha) \geq k(n-1) \tag{15}$$

where $H^{(n)}_{G_{A|B}\setminus\bigsqcup_{i\in[k]}\pi_i}(\alpha)$*, defined as follows:*

$$H^{(n)}_{G_{A|B}\setminus\bigsqcup_{i\in[k]}\pi_i}(\alpha) := \sum_{x\in V_A\setminus\bigsqcup_{i\in[k]}\pi_i} |\gamma^{-1}\alpha_x| + \sum_{x\in V_B\setminus\bigsqcup_{i\in[k]}\pi_i} |\,\mathrm{id}^{-1}\,\alpha_x| + \sum_{x\sim y\in V_b\setminus\bigsqcup_{i\in[k]}\pi_i} |\alpha_y^{-1}\alpha_x|. \tag{16}$$

It should be mentioned that the sums in the above proposition are over $\beta \setminus \bigsqcup_{i\in[k]}\pi_i$ for $\beta \in \{V_A, V_B, V_b\}$, which make up the set of the different boundary and bulk regions when removing all the different edges and vertices that will contribute in different paths $\pi_i \in \tilde{\mathcal{P}}(G_{A|B})$ in $G_{A|B}$.

Proof. Consider a set of edge disjoint paths $\{\pi_i\}_{i\in[k]} \in \tilde{\mathcal{P}}(G_{A|B})$. We can fix a path π_i for a given $i \in [k]$ where

$$\pi_i : \mathrm{id} \to \alpha_{x_1} \to \alpha_{x_2} \to \cdots \to \alpha_{x_n} \to \gamma$$

is a path that starts from id, explores $\{x_i\}_{i\in[n]}$ vertices, and ends in γ. Using Equation (14) with the path defined above, we obtain

$$H^{(n)}_{G_{A|B}}(\alpha) = |\alpha_{x_1}| + \sum_{i=1}^{n-1} |\alpha_{x_i}^{-1}\alpha_{x_{i+1}}| + |\alpha_{x_n}^{-1}\gamma| + H^{(n)}_{G_{A|B}\setminus\pi_i}(\alpha) \geq n - 1 + H^{(n)}_{G_{A|B}\setminus\pi_i}(\alpha),$$

where we have used the triangle inequality for the Cayley distance together with the fact that $|\gamma| = n - 1$. The Hamiltonian $H^{(n)}_{G_{A|B}\setminus\pi_i}(\alpha)$ is the contribution when the path π_i from $G_{A|B}$ is used.

By iteration over all the edge-disjoint paths $\{\pi_i\}_{i\in[k]} \in \tilde{\mathcal{P}}(G_{A|B})$, the desired result is obtained. The second inequality is obtained by observing that $H^{(n)}_{G_{A|B}\setminus\pi_i}(\alpha) \geq 0$, which ends the proof of this proposition. □

Proposition 4. *Given a graph G, there exist a tuple of permutations α such that $H^{(n)}_{G_{A|B}}(\alpha) = \mathrm{maxflow}(G_{A|B})$.*

Proof. By the celebrated max-flow min-cut theorem, the maximum flow in the network is equal to its minimal cut. Recall that a *cut* of a network is a partition of its set of vertices into two subsets $S \ni s$ and $T \ni t$, with the size of the cut being the number of $S - T$ edges. In the current setting, the max-flow min-cut theorem (see, e.g., [36], Theorem 8.6) implies that there exists a partition of the vertex set $\tilde{V}$ of $G_{A|B}$ (see Definition 4) into two subsets $\tilde{V} = S \sqcup T$ with $\mathrm{id} \in S$ and $\gamma \in T$, such that

$$\mathrm{maxflow}(G_{A|B}) = \Big|\{(x,y) \in \tilde{E} \,:\, x \in S \text{ and } y \in T\}\Big|.$$

For $x \in V$, we define

$$\alpha_x = \begin{cases} \mathrm{id} & \text{if } x \in S \\ \gamma & \text{if } x \in T. \end{cases}$$

Because $\mathrm{id} \in S$ and $\gamma \in T$, we have

$$\begin{aligned} H^{(n)}_{G_{A|B}}(\alpha) &= \sum_{x\in V_A} |\gamma^{-1}\alpha_x| + \sum_{x\in V_B} |\,\mathrm{id}^{-1}\,\alpha_x| + \sum_{(x,y)\in V_b} |\alpha_x^{-1}\alpha_y| \\ &= \sum_{\substack{x\in V_A \\ x\in S}} |\gamma^{-1}\alpha_x| + \sum_{\substack{x\in V_B \\ x\in T}} |\,\mathrm{id}^{-1}\,\alpha_x| + \sum_{\substack{(x,y)\in V_b \\ x\in S \text{ and } y\in T}} |\alpha_x^{-1}\alpha_y| \\ &= (n-1)\Big[|\{(x,\cdot) \in A \,:\, x \in S\}| + |\{(x,\cdot) \in B \,:\, x \in T\}| + \\ &\qquad\qquad |\{(x,y) \in E_b \,:\, x \in S \text{ and } y \in T\}|\Big] \\ &= (n-1)\,\mathrm{maxflow}(G_{A|B}), \end{aligned}$$

where in the last claim we have used the fact that there are no edges between id and γ in $\tilde{E}$; see Figure 5. □

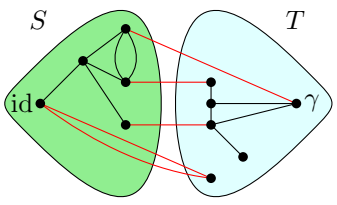


Figure 5. The different edges corresponding to a cut $S - T$ in the network $G_{A|B}$.

Proposition 5. *For all $n \geq 1$, it holds that*

$$\min_{\alpha \in \mathcal{S}_n^{|V|}} H^{(n)}_{G_{A|B}}(\alpha) = (n-1)\,\mathrm{maxflow}(G_{A|B}).$$

Proof. This proof follows from the two previous propositions. □

When all the augmenting paths in the network $G_{A|B}$ achieving the maximal flow have been identified and removed, a *clustered graph* $G^c_{A|B}$ is obtained by identifying different remaining connected permutations.

The following example provides an illustration of the different steps described above to analyze the maximum flow problem in the case of the tensor network represented in Figure 1.

Example 2. *Figure 7 represents the network $G_{A|B}$ associated with the random tensor network from Figure 1. The vertices in the network are valued in the permutation group $\mathcal{S}_n$. The network is constructed by adding two extra vertices γ and* id *by connecting all of the half-edges in A to γ and the half-edges in B to* id*. The flow approach induces a flow from* id *to γ; here, the maximum flow in Figure 7 is* 4*, where the augmenting paths achieving it are shown in color. By removing the four edge-disjoint augmenting paths, we obtain the clustered graph $G^c_{A|B}$ in Figure 8 by identifying the remaining connected edges as a single permutation cluster (i.e.,* id *with α_{15}) to form the cluster* $[\mathrm{id}, 15]$.

Theorem 4. *In the limit $D \to \infty$, for all $n \geq 1$ we have the following:*

$$\lim_{D\to\infty} \mathbb{E}\frac{1}{D^{F(G_{A|B})}}\left[\mathrm{Tr}\left(\left(D^{F(G_{A|B})-|E_\partial|}\,\rho_A\right)^n\right)\right] = m_n$$

where m_n is the number of permutations achieving the minimum of the network Hamiltonian $G^{(n)}_{A|B}$. These numbers are the moments of a probability measure $\mu_{G_{A|B}}$ which is compactly supported on $[0, +\infty)$.

Proof. For fixed n, the convergence to m_n (the number of minimizers of the Hamiltonian $G^{(n)}_{A|B}$) follows from Proposition 1 and Proposition 5. The claim that the numbers $(m_n)_n$ are the moments of a compactly supported probability measure follows basically from Prokhorov's theorem ([37], Section 5; see also [33], Footnote 2). Indeed, we note that at fixed D, the quantity

$$\mathbb{E}\frac{1}{D^{F(G_{A|B})}}\left[\mathrm{Tr}\left(\left(D^{F(G_{A|B})-|E_\partial|}\,\rho_A\right)^n\right)\right]$$

is the n-th moment of the empirical eigenvalue distribution of the random matrix $D^{F(G_{A|B})-|E_\partial|}\,\rho_A$, restricted to a subspace of dimension $D^{F(G_{A|B})}$ containing its support (this follows from the fact that $D^{F(G_{A|B})}$ is an upper bound on the rank of ρ_A; see Equation (6)). These measures have finite second moments, meaning that the sequence (index by D) is tight. The limiting moments satisfy Carleman's condition, since $m_n \leq \mathrm{Cat}_n^{|V|}$, proving that the limit measure $\mu_{G_{A|B}}$ has compact support; recall that $\mathrm{Cat}_n \leq 4^n$ is the n-th *Catalan number* (see Appendix A). Because the matrix ρ_A is positive semi-definite, $\mu_{G_{A|B}}$ must be supported on $[0, +\infty)$. □

Remark 3. *The obtained moments are provided by a graph-dependent measure. It will be shown in the following sections that such measures can be explicitly constructed if the partial order $G^o_{A|B}$ is series-parallel; see Section 6 and Theorem 5 for more details.*

In the above, the contribution terms at a large bond dimension $D \to \infty$ of $\mathbb{E}\,\mathrm{Tr}(\rho_A^n)$ are those that minimize $H^{(n)}_{G_{A|B}}(\alpha)$. As shown in Proposition 5, the quantity $H^{(n)}_{G_{A|B}}(\alpha)$ is minimized when the maximal flow is attained in $G_{A|B}$.

For later purposes, if the moment of $\tilde{\rho}_A$ is to be analyzed, then the contribution of the normalization term of $\tilde{\rho}_A$ at large bond dimension should also be considered. Recall from Proposition 2 that the contribution of the normalization term is provided by

$$\mathbb{E}[(\mathrm{Tr}\,\rho_A)^n] = \sum_{\alpha=(\alpha_x)\in\mathcal{S}_n^{|V|}} D^{n|E|-n|E_b|-h^{(n)}_G(\alpha)},\ \forall\ n\in\mathbb{N},$$

where

$$h^{(n)}_G(\alpha) := \sum_{(x,\cdot)\in E_\partial} |\,\mathrm{id}_x^{-1}\,\alpha_x| + \sum_{(x,y)\in E_b} |\alpha_x^{-1}\alpha_y|.$$

At a large dimension $D \to \infty$, the contributed terms are provided by the one that will minimize the extended Hamiltonian $h^{(n)}_{G_{A|B}}(\alpha)$ in $G_{A|B}$:

$$h^{(n)}_{G_{A|B}}(\alpha) := \sum_{x\in V_\partial} |\,\mathrm{id}^{-1}\,\alpha_x| + \sum_{(x,y)\in V_b} |\alpha_x^{-1}\alpha_y|,$$

where the first is over all the vertices V_∂ with boundary edges and where V_b are the bulk vertices.

Proposition 6. *Let $h^{(n)}_{G_{A|B}}(\alpha)$ be the extended Hamiltonian in $G_{A|B}$; then, for all $n \geq 1$ we have*

$$\min_{\alpha\in\mathcal{S}_n^{|\tilde{V}|}} h^{(n)}_{G_{A|B}}(\alpha) = 0,$$

which is achieved by identifying all the permutations with id.

Proof. To minimize the Hamiltonian $h^{(n)}_{G_{A|B}}(\alpha)$, the same approach is followed; all half-edges A are connected to γ and all half-edges B to id. However, in $h^{(n)}_{G_{A|B}}(\alpha)$ all of the boundary terms are connected to id, meaning that no path starts from id and ends in γ. Per the bulk connectivity of G, the minimum is achieved by identifying all of the permutations with id; thus, the result follows from Proposition 5. □

Remark 4. *In Proposition 6, the Hamiltonian $h^{(n)}_{G_{A|B}}(\alpha)$ is obtained by taking $A = \emptyset$ in $H^{(n)}_G(\alpha)$ (see Equation (10)). It should be mentioned that if $B = E_\partial \setminus A = \emptyset$, then the same form of the Hamiltonian $h^{(n)}_{G_{A|B}}(\alpha)$ is obtained, where instead of all the half-edges being connected to* id*, they are all connected to γ. Therefore, it can be deduced that there are no paths that start from* id *and end at γ; hence, the minimum is* 0*, which is achieved by identifying all the permutations with γ.*

Corollary 1. *For any $A \subseteq E_\partial$ moments of the normalization term, it converges to* 1*; more precisely,*

$$\lim_{D\to\infty} \mathbb{E}\Big(\mathrm{Tr}\Big(D^{-|E_\partial|}\rho_A\Big)\Big)^n = 1.$$

Proof. By taking the average, as shown in Proposition 2, we obtain the Hamiltonian $h^{(n)}_G(\alpha)$. Per the maximal flow, the Hamiltonian is minimized by identifying all of the permutations to id; therefore, we have $F(G_{A|B}) = 0$, as shown in Proposition 6. By removing all of the augmenting paths that achieve maximal flow, the obtained residual graph is trivial, with

only two disjoint vertices γ and the identity cluster $[\mathrm{id}]$. Hence, by Theorem 4 we obtain the desired result. □

6. Moment for Ordered Series-Parallel Network

In this section, the notion of a series-parallel graph (Section 6.1) is introduced. This notion allows the moment to be computed as an explicit graph-dependent measure (Section 6.2). More precisely, it is show that with the help of free probability, the corresponding graph-dependent measure can be explicitly constructed in the case where the obtained partial order $G^o_{A|B}$ is series-parallel.

6.1. Series-Parallel Graph

In this subsection, the notion of series-parallel partial orders is introduced, which in the following subsection will allow for explicit computation of the moments as graph-dependent measures.

First, the notion of series-parallel partial order [38] is recalled and some crucial definitions are provided that will play an important role in the rest of this section.

Given two partial orders $(P_i, \leq_i)$, $i = 1, 2$, we defines their *series* (resp. *parallel*) composition as follows: the base set is $P := P_1 \sqcup P_2$ and the order relation $x \leq y$ if

- $x, y \in P_i$ and $x \leq_i y$ *or* $x \in P_1$ and $y \in P_2$ in the series case;
- $x, y \in P_i$ and $x \leq_i y$ in the parallel case.

It is more convenient to represent partial orders by their *covering graphs*, where an oriented graph $G(V, E)$ is associated with a partial order $(P, \leq)$, with $V = P$ and $x \to y \in E$ if and only if $x < y$ and there does not exist z such that $x < z < y$. Recall that $x < y$ denotes $x \leq y$ and $x \neq y$. The series and parallel composition for partial orders can be elegantly interpreted in terms of directed graphs (or networks, in this context). In what follows, the terms "partial order" and "partial order graph" are used interchangeably.

Definition 11 ([38]). *Let H_1 and H_2 be two directed graphs with their respective sources s_i and sinks t_i for $i \in \{1, 2\}$. A* series-parallel network *is a directed graph $G = (V, E)$ containing two distinct vertices $s \neq t \in V$, called the* source *and the* sink*, that can be obtained recursively from the trivial network $G_{triv} = (\{s, t\}, \{\{s, t\}\})$ using the following two operations (see Figure 6):*

- *Series concatenation: $G = H_1 \,|\mathrm{S}|\, H_2$ is obtained by identifying the sink of H_1 with the source of H_2, i.e., $t_1 = s_2$.*
- *Parallel concatenation: $G = H_1 \,|\mathrm{P}|\, H_2$ is obtained by identifying the source and the sink of H_1 and H_2, i.e., $s_1 = s_2$ and $t_1 = t_2$.*

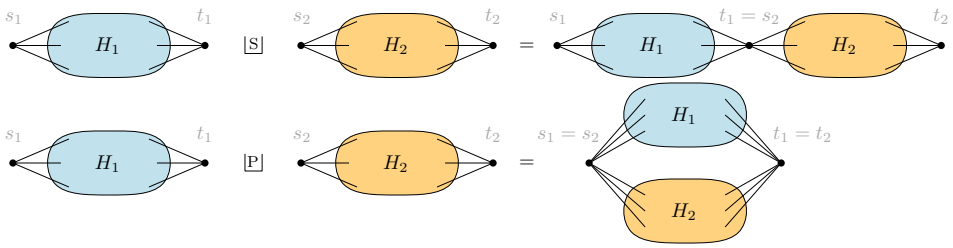


Figure 6. Series (**top**) and parallel (**bottom**) composition of two networks (graphs) $H_{1,2}$.

Remark 5. *Note that the parallel concatenation is a commutative operation, while the series concatenation is in general not commutative:*

$$\begin{aligned} \forall G, H \quad & G \,|\mathrm{P}|\, H = H \,|\mathrm{P}|\, G, \\ \textit{in general} \quad & G \,|\mathrm{S}|\, H \neq H \,|\mathrm{S}|\, G. \end{aligned}$$

From a given series-parallel network, different probability distributions can be associated that are constructed from the parallel and series concatenation introduced in Definition 11.

Definition 12. *A probability measure μ_G is associated with a series-parallel network G, defined recursively as follows:*

- *The Dirac mass at* 1 *is associated with the trivial network $G_{triv} = (\{s,t\}, \{\{s,t\}\})$:*

$$\mu_{G_{triv}} := \delta_1.$$

- *For series concatenation, the free multiplicative convolution of the parts is taken along with the measure* MP*:*

$$\mu_{G \,[\mathbb{S}]\, H} := \mu_G \boxtimes \mathrm{MP} \boxtimes \mu_H.$$

- *For parallel concatenation, the classical multiplicative convolution of the parts is taken:*

$$\mu_{G \,[\mathbb{P}]\, H} := \mu_G \times \mu_H.$$

In the above definition, the free product convolution $\boxtimes$ and the Marčhenko–Pastur distribution MP are used; refer to Appendix A for a self-contained introduction to free probability theory.

6.2. Moment as Graph-Dependent Measure

In this subsection, with the help of the series-parallel notion introduced in the previous subsection, we show that the moments m_n in Theorem 4 are explicitly constructed from a graph-dependent measure in the case where the obtained partial order $G^o_{A|B}$ is series-parallel.

Before presenting the results of this subsection, the different results obtained in the previous sections are first recalled. For a given random tensor network, as represented for example in Figure 1, the moment for a normalized quantum state corresponding to a given subregion $A \subseteq E_\partial$ of the graph was computed in Section 4 (see Propositions 1 and 2). The evaluation of the moment as $D \to \infty$ was approached via the maximal flow method, as analyzed in Section 5. The network $G_{A|B}$ was constructed from the graph G by connecting each of the regions A and B to γ and id, respectively, where the flow consists of analyzing the different paths starting from id and ending in γ. By removing all of the different augmenting paths achieving maximal flow, a clustered graph $G^c_{A|B}$ remains that is obtained by identifying different edge-connected permutations, as represented in Figure 8 for the clustered graph associated with the network $G_{A|B}$ in Figure 1. Using the maximal flow, Proposition 5 allows us to show the convergence of the moments provided by a graph-dependent measure $\mu_{G_{A|B}}$, as stated in Theorem 4. Moreover, from Proposition 6 it is deduced in Corollary 1 that the normalization terms converge to 1.

From the clustered graph $G^c_{A|B}$, a *partial order* $G^o_{A|B}$ is constructed in which the vertices in $G^o_{A|B}$ correspond to the different permutation clusters. Figure 9 illustrates the obtained partial order $G^o_{A|B}$ for the network $G_{A|B}$ shown in Figure 1. If the partial order $G^o_{A|B}$ is series-parallel (see Definition 11), convergence in moments of $\tilde{\rho}_A$ to an explicit partial order measure $\mu_{G^o_{A|B}}$ occurs, which will be explicitly shown in the following subsections.

The following theorem shows the convergence to a moment-dependent measure $\mu_{G^o_{A|B}}$ in the case where the obtained partial order $G^o_{A|B}$ is series-parallel.

Theorem 5. *For any $A \subseteq E_\partial$, assuming that the partial order $G^o_{A|B}$ is series-parallel, the limit measure from Theorem 4 can be explicitly constructed from the partial order*

$$\mu_{G_{A|B}} = \mu_{G^o_{A|B}}.$$

In particular, the moments of the reduced tensor network matrix are provided by

$$\lim_{D\to\infty} \frac{1}{D^{F(G_{A|B})}} \mathbb{E}\,\mathrm{Tr}\Big(\Big(D^{F(G_{A|B})-|E_\partial|}\rho_A\Big)^n\Big) = \int t^n \,\mathrm{d}\mu_{G^o_{A|B}}. \tag{17}$$

Proof. It suffices to show that the numbers $m_{n,G^o_{A|B}}$ are the moments of the probability measure $\mu_{G^o_{A|B}}$. This will be proved using the recursive structure of the series-parallel networks (see Definition 11) together with that of the probability measure $\mu_{G^o_{A|B}}$ (see Definition 12).

If the partial order $G^o_{A|B}$ is trivial, then it consists only of two connected components, namely, that of the identity (source) $[\mathrm{id}]$ and the sink $[\gamma]$. Hence, all the permutations associated with the connected components are fixed to be either id or γ. Thus, $m_{n,G^o_{A|B}} = 1$ for all $n \geq 1$, which are the moments of the measure $\mu_{G^o_{A|B}} = \delta_1$. In this way, the claim is seen to hold for the initial case of a trivial network.

If the partial order $G^o_{A|B}$ is the *parallel* concatenation of two networks $G^o_{A|B} = H_1 \,\boxed{\mathrm{P}}\, H_2$ having the same source and sink as $G^o_{A|B}$, then the geodesic equalities for $G^o_{A|B}$ are the disjoint union of the geodesic equalities for the vertices in H_1 and those for the vertices of H_2. In turn, this implies that

$$m_{n,G^o_{A|B}} = m_{n,H_1} \cdot m_{n,H_2}$$

for all $n \geq 1$, as there is no geodesic inequality mixing vertices from H_1 with vertices in H_2. Hence, by the induction hypothesis, we have

$$m_{n,G^o_{A|B}} = \int t^n \,\mathrm{d}\mu_{H_1} \cdot \int t^n \,\mathrm{d}\mu_{H_2} = \int t^n \,\mathrm{d}\big[\mu_{H_1} \,\boxed{\mathrm{P}}\, \mu_{H_2}\big] = \int t^n \,\mathrm{d}\mu_{G^o_{A|B}},$$

proving the claim for the parallel concatenation of networks.

Finally, the case where the network is the *series* concatenation of two networks $G^o_{A|B} = H_1 \,\boxed{\mathrm{S}}\, H_2$ is considered. This means that there is a connected component, say $[\beta]$, which is common to the two networks, as it is the sink of H_1 and the source of H_2. All geodesic equality conditions for H_1 are of the form

$$\mathrm{id} \to \alpha^{(1)}_1 \to \cdots \to \alpha^{(1)}_{k_1} \to \beta,$$

while those of H_2 are of the form

$$\beta \to \alpha^{(2)}_1 \to \cdots \to \alpha^{(2)}_{k_2} \to \gamma.$$

In particular, the geodesic equality conditions for $G^o_{A|B} = H_1 \,\boxed{\mathrm{S}}\, H_2$ are of the form

$$\mathrm{id} \to \alpha^{(1)}_1 \to \cdots \to \alpha^{(1)}_{k_1} \to \beta \to \alpha^{(2)}_1 \to \cdots \to \alpha^{(2)}_{k_2} \to \gamma.$$

The variable β is a non-constrained non-crossing partition of $[n]$, and summing over it corresponds to taking the free multiplicative convolution with respect to the Marčhenko–Pastur distribution:

$$m_{n,G^o_{A|B}} = \sum_{\substack{\beta\in \mathrm{NC}(n) \\ \alpha^{(1)}_i \leq \beta \leq \alpha^{(2)}_j}} 1 = \int t^n \,\mathrm{d}\mu_{H_1} \boxtimes \mathrm{MP} \boxtimes \mu_{H_2} = \int t^n \,\mathrm{d}\mu_{H_1 \,\boxed{\mathrm{S}}\, H_2} = \int t^n \,\mathrm{d}\mu_{G^o_{A|B}}.$$

This proves the final claim and concludes the proof. □

Example 3. *As an example, the graph illustrated in Figure 1 is considered. As described in the previous sections, the dominant terms of moments in Proposition 1 are obtained by analyzing the maximum flow in $G_{A|B}$, as shown in Figure 7, where* $\mathrm{maxflow}(G_{A|B}) = 4$. *The partial order $G^o_{A|B}$ obtained by removing from $G_{A|B}$ those edges that participate in the maximum flow, is depicted in Figure 8. Using the four augmenting paths (displayed by the colors in Figure 7), the* partial order *on the connected components of the partial order is constructed, as depicted in Figure 9.*

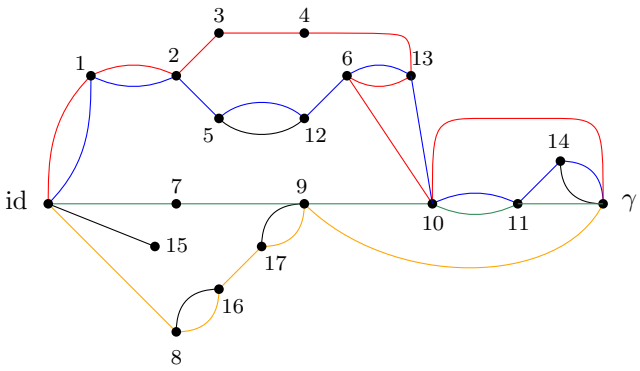


Figure 7. The network associated with the random tensor network marginal from Figure 1. The maximum flow of this network (with source id and sink γ) is 4. The four augmenting paths achieving this value are shown in color.

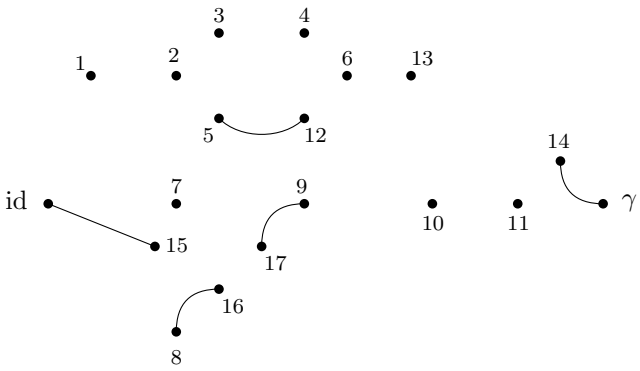


Figure 8. The clustered network corresponding to Figure 7 obtained by removing the four edge-disjoint augmenting paths.

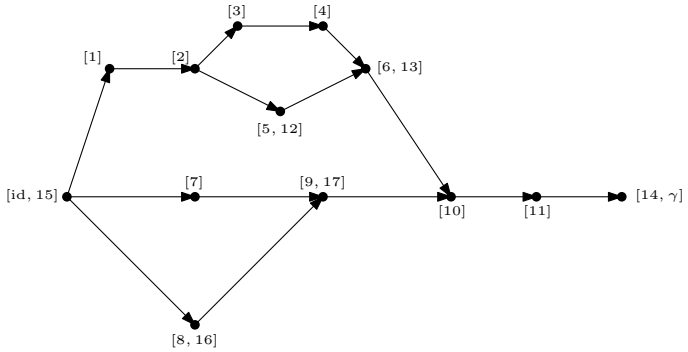


Figure 9. The order graph corresponding to the networks from Figure 1. The partial order relations are to be read from left to right. The elements of this partial-order relation are the connected components of the clustered network from Figure 8, which are eventually identified after taking into account the inequalities from the augmenting paths from Figure 7.

This process is fundamental to the presented approach; the details for one of these geodesics are provided next. For example, consider the augmenting path

$$\mathrm{id} \to 1 \to 2 \to 3 \to 4 \to 13 \to 6 \to 10 \to \gamma$$

depicted in red in Figure 7. Because the respective pair of points $(\mathrm{id}, 15)$ *and* $14, \gamma$ *in the clustered graph from Figure 8 are in the same connected components (clusters), this augmenting path gives rise to the following list of partial order relations:*

$$[\mathrm{id}, 15] \preceq [1] \preceq [2] \preceq [3] \preceq [4] \preceq [13] \preceq [6] \preceq [10] \preceq [14, \gamma].$$

The other three augmenting paths, respectively depicted in blue, green, and orange in Figure 7, give rise to the following list of inequalities:

$$\begin{aligned}
&[\mathrm{id}, 15] \preceq [1] \preceq [2] \preceq [5, 12] \preceq [6] \preceq [13] \preceq [10] \preceq [11] \preceq [14, \gamma] \\
&[\mathrm{id}, 15] \preceq [7] \preceq [9, 17] \preceq [10] \preceq [11] \preceq [14, \gamma] \\
&[\mathrm{id}, 15] \preceq [8, 16] \preceq [9, 17] \preceq [14, \gamma].
\end{aligned}$$

The partial order depicted in Figure 9 is compiled from the set of inequalities coming from the (fixed) list of augmenting paths yielding the maximum flow (here, 4). It is important to note that some connected components (clusters) can be identified in this partial order due to the anti-symmetry *property* $x \preceq y$ *and* $y \preceq x \implies x = y$*; in our example, this happens for the clusters* $[6]$ *and* $[13]$.

As an application of Theorem 5, it is possible to provide explicit moments of the measure $\mu_{G^o_{A|B},A}$*:*

$$\lim_{D\to\infty} \mathbb{E} D^{-4} \operatorname{Tr}\left[(D^{-6}\rho_A)^n\right] = m_{n,G^o_{A|B}} = \int x^n \, \mathrm{d}\mu_{G^o_{A|B}}. \tag{18}$$

The powers of D *in the normalization follow from* $|E_\partial| = 10$ *(see the boundary edges in Figure 1) and from* $\mathrm{maxflow}(G_{A|B}) = 4$*. The resulting probability measure* $\mu_{G^o_{A|B}}$ *associated with the partial order from Figure 9 is provided by*

$$\mu_{G^o_{A|B}} = \left\{\left[\mathrm{MP}^{\boxtimes 3} \boxtimes (\mathrm{MP}^{\boxtimes 2} \times \mathrm{MP})\right] \times [(\mathrm{MP} \times \mathrm{MP}) \boxtimes \mathrm{MP}]\right\} \boxtimes \mathrm{MP}^{\boxtimes 2}. \tag{19}$$

The above measure is obtained by the iterative procedure from Definition 12 as follows. First, observe that the graph in Figure 9 can be decomposed as a series *composition of three graphs* $G_1 \,|\mathrm{S}|\, G_2 \,|\mathrm{S}|\, G3$*; hence, using* $\mu_{G_2} = \mu_{G_3} = \delta_1$*, it follows that (see Figure 10)*

$$\mu_{G^o_{A|B}} = \mu_{G_1} \boxtimes \mathrm{MP} \boxtimes \mu_{G_2} \boxtimes \mathrm{MP} \boxtimes \mu_{G_3} = \mu_{G_1} \boxtimes \mathrm{MP}^{\boxtimes 2}.$$

Now, we can observe that G_1 *is a* parallel *composition of two other graphs; hence (see Figure 11)*

$$\mu_{G_1} = \mu_{G_4} \times \mu_{G_5}.$$

The graphs G_4 *and* G_5 *are now analyzed separately. First,* G_4 *can be decomposed as a series composition between the parallel composition of* G_6 *and* G_7*, and* G_8*, that is (see Figure 12)*

$$G_4 = \left(G_6 \,|\mathrm{P}|\, G_7\right) \,|\mathrm{S}|\, G_8 \implies \mu_{G_4} = (\mu_{G_6} \times \mu_{G_7}) \boxtimes \mathrm{MP} \boxtimes \mu_{G_8}.$$

Now, G_6 *and* G_7 *are series compositions of two trivial graphs; thus,* $\mu_{G_6} = \mu_{G_7} = \mathrm{MP}$ *and* $\mu_{G_8} = \delta_1$*. It follows that*

$$\mu_{G_4} = (\mathrm{MP} \times \mathrm{MP}) \boxtimes \mathrm{MP}.$$

The graph G_5 *can be decomposed as follows (see Figure 13)*

$$G_5 = G_9 \,|\mathrm{S}|\, G_{10} \,|\mathrm{S}|\left(G_{11} \,|\mathrm{P}|\, G_{12}\right) \,|\mathrm{S}|\, G_{13}.$$

In terms of the associated probability measures, it is found that

$$\mu_{G_5} = \mu_{G_9} \boxtimes \mathrm{MP} \boxtimes \mu_{G_{10}} \boxtimes \mathrm{MP} \boxtimes \left(\mu_{G_{11}} \times \mu_{G_{12}}\right) \boxtimes \mathrm{MP} \boxtimes \mu_{G_{13}}.$$

By iteratively applying series compositions, we obtain

$$\mu_{G_{11}} = \mathrm{MP}^{\boxtimes 2} \qquad \textit{and} \qquad \mu_{G_{12}} = \mathrm{MP}.$$

Thus,

$$\mu_{G_5} = \mathrm{MP}^{\boxtimes 3} \boxtimes \left(\mathrm{MP}^{\boxtimes 2} \times \mathrm{MP}\right).$$

With all of these ingredients, the specified formula for $\mu_{G^o_{A|B}}$ *is obtained.*

Remark 6. *In the example of the tensor network represented in Figure 1, the moments are computed from the factorized series-parallel structure using the flow approach. It should be mentioned that if the minimal cut approach is taken, then minimal intersecting cuts in the network exist, as shown in Figure 14. Therefore, the correction terms of the entropy can be computed as the moment of a given measure without the need for any minimal cut assumptions such as those considered in [3].*

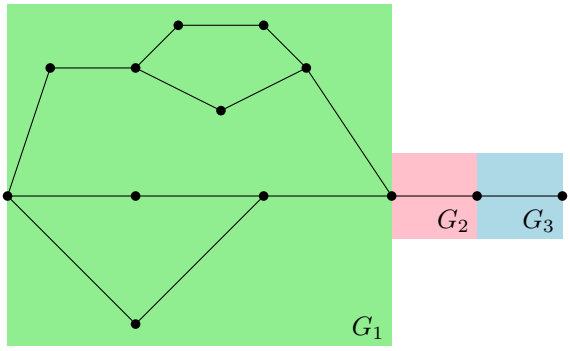


Figure 10. Representation of the partial order $G^o_{A|B}$ factorizes to a series combination of graphs G_1, G_2, and G_3.

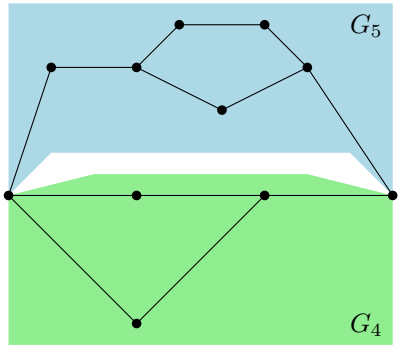


Figure 11. The graph G_1 is factorized to the parallel composition of G_4 and G_5.

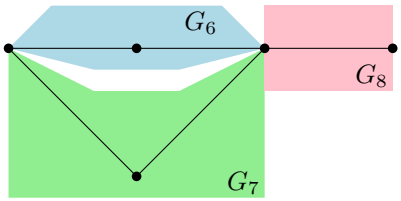


Figure 12. The graph G_4 is factorized to the series composition of the graph G_8, with graph G_6 parallel to G_7.

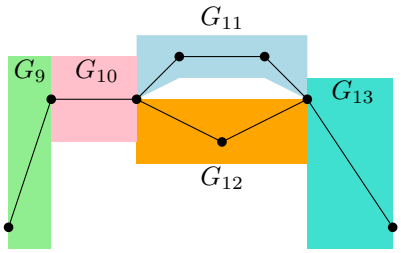


Figure 13. The graph G_5 factorizes as a series composition of G_9 and G_{10}, with G_{11} composed in parallel to G_{12} and in series with G_{13}.

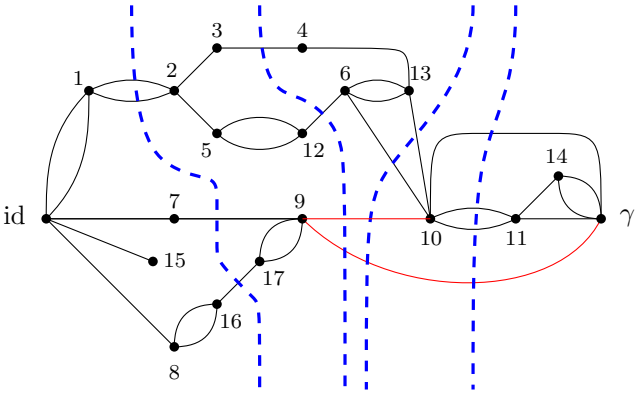


Figure 14. Different cuts of size 4 in the network of Figure 7 are represented with blue dashed lines. The maximum flow in the network is 4. Note that some of these minimal cuts share some edges, which are represented in red.

Remark 7. *The obtained measure $\mu_{G^o_{A|B}}$ for a given ordered series-parallel graph $G^o_{A|B}$ has compact support where it combines the Marčhenko–Pastur distribution with the classical product measure and free product convolution constructed from the structure of $G^o_{A|B}$.*

7. Examples of Series-Parallel Networks

In this section, the results obtained previously are applied to various random tensor networks that possess an induced series-parallel order. The discussion begins with simple cases and progresses towards more physically relevant cases.

7.1. Single Vertex Network

The simplest possible case is considered first, that is, a tensor network having only one vertex, no bulk edges, and two boundary half-edges; see Figure 15. For this network, the associated random tensor

$$\Psi_G \in \mathbb{C}^D \otimes \mathbb{C}^D$$

has i.i.d. standard complex Gaussian entries. The two boundary half-edges are partitioned into two one-element sets, A and $B = \bar{A}$. From this tensor, the reduced matrix is constructed as

$$\rho_A := \mathrm{Tr}_B |\Psi_G\rangle\langle\Psi_G|$$

by partial tracing of the half-edge B. Note that in this very simple case the matrix ρ_A can also be seen as a product of the matricization of the tensor Ψ_G with its Hermitian adjoint; hence, ρ_A is a Wishart random matrix (see Appendix A for the definition and basic properties of Wishart matrices).

To analyze the large D spectral properties of ρ_A, the network $G_{A|B}$ is first constructed by connecting all the half-edges in G that belong to A to a new vertex γ and connecting those in B to a new vertex id. The flow analysis of this network is trivial; a unique path

from id to γ exists, meaning that the maximum flow is 1 and the residual network is empty (both edges in the network are used for the construction of the unique maximum flow).

Because there is a unique path achieving maximum flow and only a single vertex in the network, the partial order induced by the path is very simple: $\mathrm{id} - \alpha_1 - \gamma$. Hence, the only condition on the permutation $\alpha_1 \in \mathcal{S}_n$ is that it should lie on the geodesic between the identity permutation id and the full cycle permutation $\gamma \in \mathcal{S}_n$. Thus, we obtain a series network; see the bottom right-hand diagram in Figure 15. The limit moment distribution is MP, that is, the Marčhenko–Pastur distribution (of parameter 1). This matches previously obtained results about the *induced measure* of mixed quantum states (density matrices) [39–41]. Indeed, the matrix ρ_A can be interpreted in quantum information theory as the partial trace of the rank-one matrix in the direction of a random Gaussian vector $\Psi_G \in \mathbb{C}^d \otimes \mathbb{C}^d$. Up to normalization, this random density matrix belongs to the ensemble of induced density matrices. The fact that the two factors of the tensor product have equal dimensions corresponds to taking the *uniform measure* on the (convex, compact) set of density matrices [42,43]. The statistics of the eigenvalues of such random matrices have been extensively studied in the literature. In particular, the asymptotic von Neumann entropy has been studied by Page [44–46], who conjectured that

$$\mathbb{E}S(\rho_A) = \sum_{i=D+1}^{2D} \frac{1}{i} - \frac{D-1}{2D} \sim \log D - \frac{1}{2} \quad \text{as } D \to \infty.$$

The reader is referred to Section 8 for a derivation of such statistics in the context of this work.

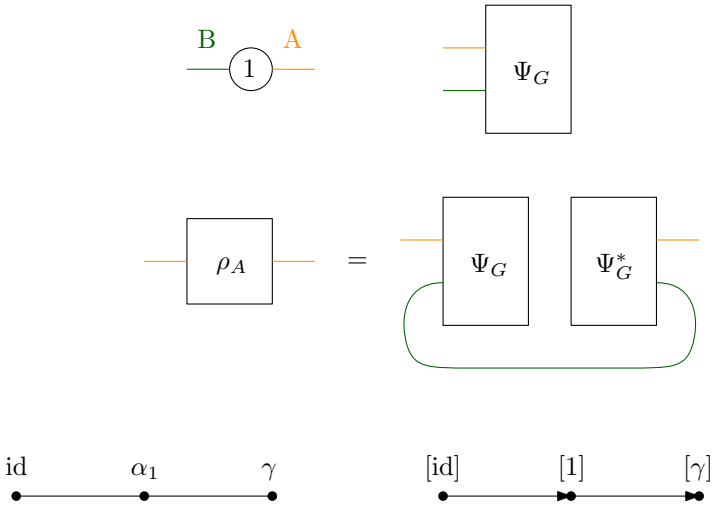


Figure 15. A single vertex network: the top row shows the network and the associated random tensor Ψ_G, the middle row shows the reduced matrix ρ_A obtained by partial tracing of the edge B between Ψ_G and Ψ_G^*, and the bottom row shows the network $G_{A|B}$ and partial-order graph $G^o_{A|B}$.

7.2. Series Network

We now consider a tensor network consisting of s vertices arranged in a path graph with two half-edges at the end points. This network is depicted along with the various steps needed to compute the limiting spectrum distribution of the reduced matrix. The network associated with the graph (where the partition of the half-edges is clear) has a single path from the source to the sink, meaning that the maximum flow is unity, see Figure 16.

The residual graph obtained by removing the edges from the unique path achieving maximum flow is empty; hence, the partial order on the vertices is again a total order:

$$\mathrm{id} \preceq \alpha_1 \preceq \cdots \preceq \alpha_s \preceq \gamma.$$

Thus, a *series network* is obtained and the final measure is obtained by applying s times the series concatenation procedure from Definition 12, yielding

$$\mu_{G^o_{A|B}} = \underbrace{\mathrm{MP} \boxtimes \mathrm{MP} \boxtimes \cdots \boxtimes \mathrm{MP}}_{s \text{ times}} = \mathrm{MP}^{\boxtimes s}.$$

Very similar results were previously obtained by Cécilia Lancien [47]; see also [25]. This measure is commonly known as the *Fuss–Catalan* distribution of order s [48]; see also Theorem A5. In combinatorics, its moments are known as the Fuss–Catalan numbers

$$\int t^n \, \mathrm{dMP}^{\boxtimes s}(t) = \frac{1}{sn+1}\binom{sn+n}{n}$$

and its entropy ([49], Proposition 6.2) is

$$\int -t \log t \, \mathrm{dMP}^{\boxtimes s}(t) = \sum_{i=2}^{s+1} \frac{1}{i}.$$

Such tensor network states have already been considered in quantum information theory [43,49,50].

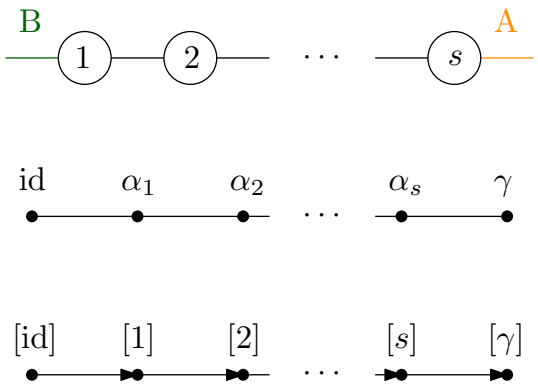


Figure 16. A series network; the s vertices of the network are arranged in a line, with two half-edges at the end points. The maximum flow is 1, and is unique. The single path realizing the flow induces a (total) order $\alpha_1 \leq \alpha_2 \leq \cdots \leq \alpha_s$ on the geodesic permutations.

7.3. Two-Dimensional Lattice

We now discuss a physically relevant network consisting of a rectangle that is part of a $2D$ lattice (part of $\mathbb{Z}^2$). Two integer parameters are considered, namely, the length L and height H of the rectangle, resulting in $H \cdot L$ vertices. The vertices are connected by the edges inherited from the $\mathbb{Z}^2$ lattice; see Figure 17. The left-most (resp. right-most) columns of vertices are assigned half-edges that belong to the class B (resp. A) of the half-edge partition defining the two regions.

The flow network corresponding to the graph and the partition $A|B$ is depicted in the top diagram in Figure 18. The maximum flow in this network is H; one can consider H parallel horizontal paths which go from id to γ. Note that the set of H edge-disjoint paths in the network achieving the maximum flow is unique. In this case, the residual network is non-empty, with H clusters of the form

$$\mathcal{C}_j := \{[i, j] \, : \, i = 1, \ldots, L\}.$$

The order relation on the clusters is again a total order on L points; see the bottom diagram in Figure 18. Again we obtain the Fuss–Catalan distribution:

$$\mu_{G^o_{A|B}} = \mathrm{MP}^{\boxtimes L}.$$

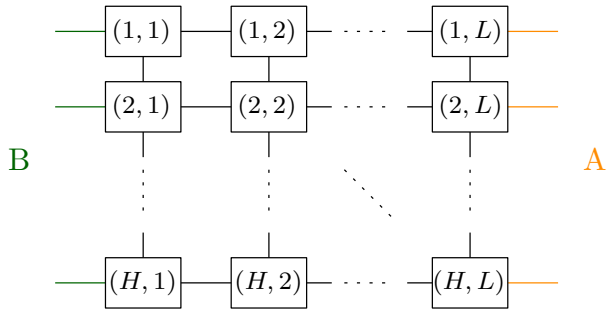


Figure 17. A network corresponding to a $H \times L$ 2D lattice. Half-edges are attached to vertices on the left and right boundaries, corresponding to a tensor in $(\mathbb{C}^D)^{\otimes 2H}$.

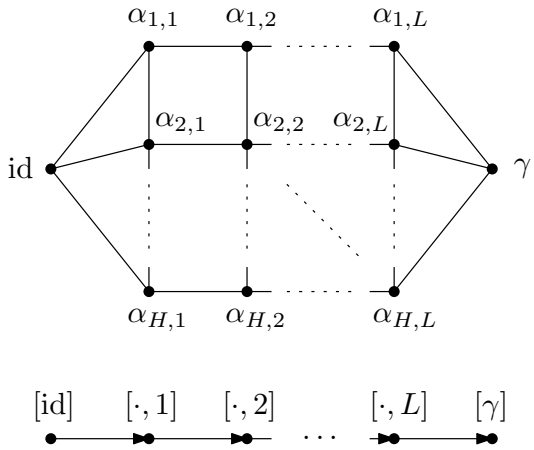


Figure 18. The network associated with the $H \times L$ $\mathbb{Z}^2$ lattice fragment. The maximum flow in this network is H, corresponding to horizontal edge-disjoint paths. These H paths induce a total order on the L vertex clusters, each having H vertices.

8. Results for Normalized Tensor Network States

This section presents the main technical contribution of this paper. With the help of all the results obtained from the previous sections, the Rényi and von Neumann entropies for a given approximated normalized state $\tilde{\rho}_A := D^{-|E_\partial|}\rho_A$ associated with a given boundary subregion $A \subseteq E_\partial$ can be computed. The main results of this section first consist in showing the weak convergence of moments associated with an approximated reduced state $\tilde{\rho}_A$ that is associated with a given boundary region A in Theorem 7. Moreover, Corollary 2 shows the existence of correction terms as moments of a graph-dependent measure, which can be explicitly computed in the case of an obtained series-parallel partial order $G^o_{A|B}$.

In Section 8.1, different concentration inequalities will be shown, which in Section 8.2 will allow the main results of this section to be presented.

8.1. Concentration

In this subsection, we provide different concentration results which will allow the main technical contribution to be presented in the following subsection.

First, we recall the following theorem which estimates the deviation probability of polynomials in Gaussian random variables. This theorem is relevant for different concentration results that will be proved later in this subsection.

Theorem 6 ([51], Theorem 6.7)**.** *Let g be a polynomial in m variables of degree q; then, if $X_1, \cdots, X_m$ are independent centered Gaussian random variables,*

$$\forall t > 0, \quad \mathbb{P}\Big(\big|g(X_1, \cdots, X_m) - \mathbb{E}\, g\big| > t(\mathrm{Var}(g))^{\frac{1}{2}}\Big) \leq \exp\Big\{\Big(-c_q\, t^{\frac{2}{q}}\Big)\Big\},$$

where $V(g)$ is the variance of $g(X_1, \cdots, X_m)$ and c_q is a constant which depends only on q.

Proposition 7. *Let G be a bulk connected graph and let $A \subseteq E_\partial$; then,*

$$\mathbb{P}\Big(\Big|\operatorname{Tr}\tilde{\rho}_A - 1\Big| > \varepsilon\Big) \leq \exp\Big\{\Big(-c_{|E|}\varepsilon^{\frac{1}{|E|}} D^{\frac{|E_b|}{2|E|}}\Big)\Big\},$$

where $\tilde{\rho}_A := D^{-|E_\partial|}\rho_A$.

Proof. First, it should be noted that $\operatorname{Tr}\rho_A$ is a polynomial in $|g_x\rangle \in \mathcal{H}_x$ of total degree $2|E|$. Moreover, recall that for a random Gaussian vector $|g_x\rangle \in \mathcal{H}_x$, the following holds:

$$\forall x \in V, \quad \mathbb{E}[|g_x\rangle\langle g_x|] = \mathrm{id}_x \quad \text{and} \quad \mathbb{E}\Big[|g_x\rangle\langle g_x|^{\otimes 2}\Big] = \mathrm{id}_x + F_x$$

where id_x and F_x act on all the edges of the Hilbert space to generate the local Hilbert space for each vertex x. Moreover, it is implicitly assumed that $\mathrm{id}_x \equiv \mathrm{id}_x^{\otimes 2}$ and that the Swap operator F_x is a unitary representation of a permutation element in $\mathcal{S}_2$.

It is easy to check that the variance $\mathrm{Var}(\operatorname{Tr}\tilde{\rho}_A)$ provides

$$\mathrm{Var}(\operatorname{Tr}\tilde{\rho}_A) = \mathbb{E}\Big[(\operatorname{Tr}\tilde{\rho}_A)^2\Big] - (\mathbb{E}[\operatorname{Tr}(\tilde{\rho}_A)])^2 = O\Big(D^{-|E_b|}\Big),$$

where we have used the fact that

$$\begin{aligned}
\mathbb{E}\Big[(\operatorname{Tr}\tilde{\rho}_A)^2\Big] &= \operatorname{Tr}\left(\bigotimes_{e\in E_b} |\Omega_e\rangle\langle\Omega_e|^{\otimes 2} \bigotimes_{x\in V} \mathbb{E}\Big[|g_x\rangle\langle g_x|^{\otimes 2}\Big]\right) \\
&= 1 + D^{|E_\partial|} \sum_{\alpha_e \in \mathcal{S}_2} \prod_{e\in E_b} \operatorname{Tr}\Big(|\Omega_e\rangle\langle\Omega_e|^{\otimes 2} U_{\alpha_e}\Big) \prod_{e\in E_\partial} \operatorname{Tr}(U_{\alpha_e}) \\
&= 1 + O\Big(D^{-|E_b|}\Big).
\end{aligned}$$

In the last equality, the bulk contribution is of $D^{-|E_b|}$, while the boundary edges contribute with $D^{|E_\partial|}$. The second term of the variance is

$$\mathbb{E}[\operatorname{Tr}(\tilde{\rho}_A)] = \operatorname{Tr}\left(\bigotimes_{e\in E_b} |\Omega_e\rangle\langle\Omega_e| \bigotimes_{x\in V} \mathbb{E}[|g_x\rangle\langle g_x|]\right) = 1.$$

By combining the variance $\mathrm{Var}(\operatorname{Tr}(\tilde{\rho}_A))$ with Proposition 6, we have

$$\mathbb{P}\Big(\big|\operatorname{Tr}\tilde{\rho}_A - \mathbb{E}\operatorname{Tr}\tilde{\rho}_A\big| > \varepsilon\Big) \leq \exp\Big\{\Big(-c_{|E|}\varepsilon^{\frac{1}{|E|}} D^{\frac{|E_b|}{2|E|}}\Big)\Big\},$$

where $\varepsilon := t\Big(D^{-\frac{|E_b|}{2}}\Big)$ with $c_{|E|} > 0$ is a constant that depends only on the total number of edges $|E|$. □

Proposition 8. *Let G be a bulk connected graph and let $A \subseteq E_\partial$; then, the following holds:*

$$\forall n > 1, \quad \mathbb{P}\Big(\Big|\frac{1}{D^{F(G_{A|B})}}\operatorname{Tr}(\sigma_A^n) - \frac{1}{D^{F(G_{A|B})}}\mathbb{E}[\operatorname{Tr}(\sigma_A^n)]\Big| > \varepsilon\Big) \leq \exp\Big\{\Big(-c_{2n|E|} D^{\frac{1}{2n|E|}} \varepsilon^{\frac{1}{n|E|}}\Big)\Big\},$$

where $\sigma_A := D^{F(G_{A|B})}\tilde{\rho}_A$.

Proof. The proof of this proposition follows the same spirit as the proof of the proposition above. We remark that $\operatorname{Tr}\sigma_A^n$ is a $2n|E|$ polynomial in $|g_x\rangle$. Moreover the variance was estimated in [33], Lemma 14, where

$$\operatorname{Var}\left(\frac{1}{D^{F(G_{A|B})}}\operatorname{Tr}(\sigma_A^n)\right) = O\left(\frac{1}{D}\right).$$

By defining $\varepsilon := tD^{-\frac{1}{2}}$, the result follows. □

8.2. Entanglement Entropy

This subsection introduces the main technical contribution of this work. With the help of the concentration results, we first assume and use the approximate normalized state $\tilde{\rho}_A := D^{-|E_\partial|}\rho_A$. It is shown that as $D \to \infty$, the average Rényi and von Neumann entanglement entropy with correction terms can be computed. In particular, if the obtained partial order is series-parallel, then the correction terms are provided as moments of a partial order-dependent measure $\mu_{G^o_{A|B}}$.

Recall from Section 3.2 that the rank of the approximate normalized state is upper-bounded by $D^{F(G_{A|B})}$. We consider the approximate normalized quantum state $\tilde{\rho}_A$ restricted to its support and its empirical measure $\mu_A^{(D)}$, defined as

$$\sigma_A := D^{F(G_{A|B})}\tilde{\rho}_A^S, \quad \text{and} \quad \mu_A^{(D)} := \frac{1}{D^{F(G_{A|B})}}\sum_{\lambda\in\operatorname{spec}(\sigma_A)}\delta_\lambda,$$

where $\tilde{\rho}_A^S$ is the reduced approximate normalized state restricted on its support. The definition of $\tilde{\rho}_A$ and the empirical measure $\mu_A^{(D)}$ allow for the demonstration in Theorem 7 of the weak convergence of $\mu_A^{(D)}$ to $\mu_{G_{A|B}}$. In particular, if the obtained partial order $G^o_{A|B}$ is series-parallel, then Theorem 5 yields weak convergence to $\mu_{G_{A|B}}$. In Corollary 2, this result allows us to compute the Rényi and von Neumann entanglement entropy.

Recall that a measure $\mu^{(D)}$ *converges weakly* to a measure μ if the following holds for any bounded continuous function $f : \mathbb{R} \to \mathbb{R}$:

$$\forall\varepsilon > 0, \quad \lim_{D\to\infty}\mathbb{P}\left(\left|\int f(t)\mathrm{d}\mu^{(D)}(t) - \int f(t)\mathrm{d}\mu(t)\right| \le \varepsilon\right) = 1.$$

Theorem 7. *Let $A \subseteq E_\partial$ be a boundary region in the graph G. The empirical measure $\mu_A^{(D)}$ associated to the approximated normalized state σ_A converges weakly to $\mu_{G_{A|B}}$. More precisely, for all bounded continuous functions $f : \mathbb{R} \to \mathbb{R}$, we have*

$$\forall\varepsilon > 0, \quad \lim_{D\to\infty}\mathbb{P}\left(\left|\int f(t)\mathrm{d}\mu_A^{(D)}(t) - \int f(t)\mathrm{d}\mu_{G_{A|B}}(t)\right| \le \varepsilon\right) = 1.$$

Proof. As shown in Theorem 4, the moments converge to a unique measure $\mu_{G_{A|B}}$. In the particular case of an ordered series-parallel graph $G^o_{A|B}$, an explicit graph dependent measure $\mu_{G^o_{A|B}}$ is obtained. Recall from Theorem 4 that

$$\frac{1}{D^{F(G_{A|B})}}\mathbb{E}\Big[\operatorname{Tr}(\sigma_A^n)\Big] \xrightarrow[D\to\infty]{} m_n = \int t^n \mathrm{d}\mu_{G_{A|B}}(t).$$

From standard probability theory results, convergence in probability implies weak convergence (see [52], Theorem 25.2). For this, it is only necessary to show the decreasing scaling of the variance as $D \to \infty$. Using ([33], Lemma 14), we have

$$\mathrm{Var}\left(\frac{1}{D^{F(G_{A|B})}}\mathrm{Tr}(\sigma_A^n)\right) = O\left(\frac{1}{D}\right), \quad (D \to \infty),$$

establishing the weak convergence of $\mu_A^{(D)}$ to $\mu_{G_{A|B}}$. In particular, if the graph is series-parallel, then we obtain $\mu_{G^o_{A|B}}$. □

Lemma 1. *Let $A \subseteq E_\partial$ be a boundary region and let $m_n^{(D)}$ be the moment associated to the empirical measure $\mu_A^{(D)}$. Then, the following holds:*

$$\mathbb{P}\left(\left|\mathbb{E}\log\left(m_n^{(D)}\right) - \log\left(\mathbb{E}m_n^{(D)}\right)\right| > \varepsilon\right) \xrightarrow[D\to\infty]{} 1 \quad \textit{where} \quad m_n^{(D)} := \frac{1}{D^{F(G_{A|B})}}\mathbb{E}\Big[\mathrm{Tr}(\sigma_A^n)\Big].$$

Proof. By Proposition 8 and Jensen's inequality, $\mathbb{E}\log\left(m_n^{(D)}\right) \leq \log\left(\mathbb{E}m_n^{(D)}\right)$ holds. It remains to show that $\mathbb{E}\log\left(m_n^{(D)}\right) \geq \log\left(\mathbb{E}m_n^{(D)}\right)$ holds with high probability. Fixing $\varepsilon > 0$, from Proposition 8 it is the case that

$$m_n^{(D)} \geq \mathbb{E}m_n^{(D)} - \delta \quad \text{with} \quad 0 < \delta \leq \frac{\varepsilon}{\varepsilon+1}\mathbb{E}m_n^{(D)}$$

holds with probability $1 - \exp\left\{\left(-c_{2n|E|}D^{\frac{1}{2n|E|}}\delta^{\frac{1}{n|E|}}\right)\right\}$. It is easy to check that the following inequalities hold:

$$\begin{aligned}\log\left(m_n^{(D)}\right) \geq \log\left(\mathbb{E}m_n^{(D)} - \delta\right) &= \log\left(\mathbb{E}m_n^{(D)}\right) + \log\left(1 - \frac{\delta}{\mathbb{E}m_n^{(D)}}\right), \\ &\geq \log\left(\mathbb{E}m_n^{(D)}\right) - \frac{\delta}{\mathbb{E}m_n^{(D)} - \delta} \geq \log\left(\mathbb{E}m_n^{(D)}\right) - \varepsilon.\end{aligned}$$

Therefore, $\mathbb{E}\log\left(m_n^{(D)}\right) \geq \log\left(\mathbb{E}m_n^{(D)}\right) - \varepsilon$ occurs with probability at least

$$1 - \exp\left\{\left(-c_{2n|E|}D^{\frac{1}{2n|E|}}\delta_{\max}^{\frac{1}{n|E|}}\right)\right\} \quad \text{where} \quad \delta_{\max} = \frac{\varepsilon}{\varepsilon+1}\mathbb{E}m_n^{(D)}.$$

As $D \to \infty$, $\mathbb{E}m_n^{(D)}$ converges; hence, $\delta_{\max} = O(1)$, showing that the probability estimate above converges to 1 and completing the proof. □

For completeness, the following proposition from [30] is recalled, which plays a key role in the proof of the main result.

Proposition 9 ([30], Proposition 4.4)**.** *Let f be a continuous function on $\mathbb{R}$ with polynomial growth and with ν_n being a sequence of probability measures that converges in moments to a compactly supported measure ν. Then, $\int f \mathrm{d}\nu_n \to \int f \mathrm{d}\nu$.*

Corollary 2. *Let $A \subseteq E_\partial$ be a boundary region in G and let $\tilde{\rho}_A$ be the approximated reduced normalized state. Then, the averaged Rényi and von Neumann entropies converge weakly as $D \to \infty$, and the following holds:*

$$F(G_{A|B}) \log D - \mathbb{E}S_n(\tilde{\rho}_A) \xrightarrow[D\to\infty]{} \frac{1}{n-1} \log\left(\int t^n \, \mathrm{d}\mu_{G_{A|B}} \right),$$
$$F(G_{A|B}) \log D - \mathbb{E}S(\tilde{\rho}_A) \xrightarrow[D\to\infty]{} \int t \, \log t \, \mathrm{d}\mu_{G_{A|B}}$$

where $F(G_{A|B}) := \mathrm{maxflow}(G_{A|B})$.

Proof. The proof of this corollary is a direct consequence of the concentration results obtained in the previous subsection and the weak convergence of $\mu_A^{(D)}$ to $\mu_{G_{A|B}}$.

First, the Rényi entropy is considered. For this, we have

$$F(G_{A|B}) \log D - \mathbb{E}S_n(\tilde{\rho}_A) = \frac{1}{1-n} \mathbb{E} \log\left(m_{n,A}^{(D)} \right), \quad \text{where} \quad m_n^{(D)} := \frac{1}{D^{F(G_{A|B})}} \mathbb{E}\Big[\mathrm{Tr}\big((\sigma_A)^n \big) \Big]$$

and recall that $\sigma_A := D^{F(G_{A|B})} \tilde{\rho}_A^S$ is restricted to the support of $\tilde{\rho}_A := D^{-|E_\partial|} \rho_A$. Using Lemma 1 and in the limit $D \to \infty$, we have

$$F(G_{A|B}) \log D - \mathbb{E}S_n(\tilde{\rho}_A) \xrightarrow[D\to\infty]{} \frac{1}{n-1} \log\left(\int t^n \, \mathrm{d}\mu_{G_{A|B}} \right).$$

For the von Neumann entropy, let $\{\lambda_i\} \in \mathrm{spec}(\sigma_A)$ and $\{\tilde{\lambda}_i\} \in \mathrm{spec}(\tilde{\rho}_A)$. It follows directly that

$$\mathbb{E}S(\tilde{\rho}_A) = -\mathbb{E} \sum_i \tilde{\lambda}_i \log(\tilde{\lambda}_i) = F(G_{A|B}) \log(D) - \frac{1}{D^{F(G_{A|B})}} \mathbb{E} \sum_i \lambda_i \log(\lambda_i).$$

Defining the function $f : \mathbb{R} \to \mathbb{R}$ as $f(t) := t \log t$, by combining Proposition 9 and Theorem 7 we obtain the following weak convergence as $D \to \infty$:

$$F(G_{A|B}) \log D - \mathbb{E}S(\tilde{\rho}_A) = \frac{1}{D^{F(G_{A|B})}} \mathbb{E}\left(\sum_i f(\lambda_i) \right) \xrightarrow[D\to\infty]{} \int f(t) \, \mathrm{d}\mu_{G_{A|B}}$$

where the measure $\mu_{G_{A|B}}$ is defined on a compact support, completing the proof of the corollary. In the particular case where the obtained poset structure $G^o_{A|B}$ is series-parallel, the obtained graph dependent measure is explicitly provided by $\mu_{G_{A|B}} = \mu_{G^o_{A|B}}$ using Theorem 5. □

9. Conclusions

Given a general graph with a boundary region and a bulk region, the main goal of this work is to compute the entanglement entropy (the Rényi and the von Neumann entropy) of a given sub-boundary region A of the graph. By analyzing the moments of a state associated with the region A as $D \to \infty$ and assisted by the (maximal) flow approach, we compute the leading terms contributing to the moment. By analyzing and removing all of the augmenting paths of the network $G_{A|B}$ constructed by connecting the region A to the total cycle γ and id to the region B starting from id and ending in γ), a cluster graph $G^c_{A|B}$ is obtained by identifying all of the remaining edge-connected permutations. The flow approach induces a natural-order poset structure, represented by the induced poset order $G^o_{A|B}$. The maximal flow approach allows for the deduction of the moment convergence to the moment of a unique graph-dependent measure $\mu_{G_{A|B}}$. This result allows for the

deduction of the higher-order correction terms of the Rényi and von Neumann entropy, provided by a graph-dependent measure $\mu_{G_{A|B}}$. Moreover, it has been shown that with the help of free probability theory, if the obtained partial order $G^o_{A|B}$ is series-parallel, then the associated graph-dependent measure $\mu_{G_{A|B}} = \mu_{G^o_{A|B}}$ can be explicitly provided, which contributes to the higher-order correction terms of each of the Rényi and von Neumann entanglement entropies.

In this work, no assumption regarding the minimal cuts is made; from the duality in the maximal flow approach, different minimal cuts can be obtained which may intersect in different edges. Moreover, the higher-order correction terms in the entanglement entropy can describe the quantum corrections beyond the area law behavior of the expected Ryu–Takayanagi entanglement entropy in the context of AdS/CFT. It has been previously argued in the literature that if higher-order correction terms in the random tensor network setting are to be considered, it is necessary to go beyond the maximally entangled state and consider general link states representing the bulk matter field. In this work, the obtained higher-order quantum fluctuation of entanglement entropy is achieved with only maximally entangled states, which we interpret as fluctuations of spacetime itself without any need for bulk fields represented by a generic link state.

Author Contributions: All authors have contributed equally to this work. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the ANR projects ESQuisses https://esquisses.math.cnrs.fr/ (accessed on 7 April 2025), grant number ANR-20-CE47-0014-01, and by STARS https://www.math.univ-toulouse.fr/~gcebron/STARS.php (accessed on 7 April 2025), grant number ANR-20-CE40-0008, as well as by the PHC program *Star* (Applications of random matrix theory and abstract harmonic analysis to quantum information theory). K.F. acknowledges support from a NanoX project grant https://nanox-toulouse.fr/ (accessed on 7 April 2025) .

Data Availability Statement: No data has been generated or used in this work.

Acknowledgments: Cécilia Lancien is thanked for sharing preliminary notes on very similar questions, and the anonymous referees are acknowledged for their valuable comments on this work.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Basics of the Combinatorial Approach to Free Probability Theory

In this section, the necessary material on *combinatorics* and *free probability* needed to understand the rest of this section is recalled and presented. All of the material introduced here is standard and can be found in [53,54].

Let $\pi := \{V_1, \cdots, V_n\}$ (not to be confused with π_i introduced in Section 5 representing the different paths) be a partition of a finite totally-ordered set S such that $\bigsqcup_{i\in[n]} V_i = S$. The sets $\{V_i\}$ are called the *blocks* of π. The notation $p \sim_\pi q$ is used if p and q belong to the same block of π. A partition π of a set S is called *crossing* if there exist $p_1 < q_1 < p_2 < q_2$ in S such that $p_1 \sim_\pi p_2 \nsim_\pi q_1 \sim q_2$. A partition is called *non-crossing* if π does not cross. The set of non-crossing partitions of S is denoted by $\mathrm{NC}(S)$. In particular, if $S = \{1, \cdots n\}$, then the non-crossing partition set is denoted by $\mathrm{NC}(n)$. The set of non-crossing partitions plays a crucial role in various areas, from combinatorics [55] to random matrices and free probability theory, which is the main focus here. Moreover, the following crucial result is noted [53]: *there exists a one-to-one correspondence between the non-crossing partition set and the*

set of permutations α *in a geodesic between* γ *and* id, *i.e.*, $|\alpha| + |\alpha^{-1}\gamma| = |\gamma|$. Another important fact is that the cardinality $|\mathrm{NC}(n)| = \mathrm{Cat}_n$, where

$$\mathrm{Cat}_n := \frac{1}{n+1}\binom{2n}{n} \tag{A1}$$

are the *Catalan numbers*. For more combinatorial details and properties of Catalan numbers and non-crossing partitions, see [53]. We assume that $(\alpha_1, \cdots, \alpha_k)$ is a k-tuple of permutations in $\mathcal{S}_n$ such that

$$|\alpha_1| + \sum_{i\in[k-1]} |\alpha_i^{-1}\alpha_{i+1}| + |\alpha_k^{-1}\gamma| = |\gamma|. \tag{A2}$$

These are geodesics between id and γ. The cardinality of the set of k-tuple permutations satisfying the geodesic Equation (A2) is known as the *Fuss–Catalan* numbers, provided by

$$\mathrm{FC}_{n,k} := \frac{1}{nk+1}\binom{n+nk}{n},$$

which generalize the Catalan numbers for $k = 1$.

The *free probability theory* tools used in this work are now introduced. The intrinsic link between free probability theory and combinatorics is also noted, and some examples are provided to illustrate it. The combinatorics presented here will be used in the rest of this section to help understand the main result.

A *non-commutative probability space* is defined as a pair $(\mathcal{A}, \omega)$, where $\mathcal{A}$ is a unital C^*-algebra with a state $\omega : \mathcal{A} \to \mathbb{C}$ such that $\omega(1_\mathcal{A}) = 1$. The elements $a \in \mathcal{A}$ are called *non-commutative variables*. In the non-commutative probability space, the distribution law μ_a can be associated to $a \in \mathcal{A}$, defined as $\mu_a = \omega(a)$.

Before providing concrete examples of non-commutative probability spaces, the notion of *freeness*, which plays a crucial role in the non-commutative probability world, is recalled. The notion of freeness generalizes "classical" independence when the algebra $\mathcal{A}$ is commutative. For n given non-commutative random variables $\{a_i\} \in \mathcal{A}$, these are said to be *free independent* if the following holds for any polynomials $\{p_i\}$:

$$\omega(a_1 a_2 \cdots a_n) = 0 \tag{A3}$$

whenever $\omega(p_k(a_{i_k})) = 0$ for $k \in [n]$ and for two non-adjacent indices i_k and i_{k+1}. With the definition of free independence, the following can be checked for two free independent variables a_1 and a_2:

$$\omega((a_1 - \omega(a_1))(a_2 - \omega(a_2))) = \omega(a_1 a_2) - \omega(a_1)\omega(a_2) = 0 \tag{A4}$$

which generalizes the notion of standard independence in the commutative setting, where $\mathbb{E}(a_1 a_2) = \mathbb{E}(a_1)\mathbb{E}(a_2)$ for two commutative random variables a_1, a_2 in a commutative probability space.

Definition A1. *Let* $(\mathcal{A}_N, \omega_N)$ *with* $N \in \mathbb{N}$ *and* $(\mathcal{A}, \omega)$ *be non-commutative probability spaces. The sequence* $a_N \in \mathcal{A}_N$ *is said to converge weakly to* $a \in \mathcal{A}$ *as* $N \to \infty$ *if the following holds:*

$$\lim_{N\to\infty} \omega_N((a_N)^n) = \omega(a^n) \quad \forall n \in \mathbb{N}, \tag{A5}$$

where $\omega(a^n) = \int x^n \mathrm{d}\mu_a(x)$ *are the moments of* a.

To illustrate concrete non-commutative probability spaces, some classical examples are provided. The first example we consider is the "classical" probability space corresponding to a commutative algebra. Let (Ω, Σ, μ), where Ω is a set, Σ is a σ-algebra, and μ is a probability measure. We define $\mathcal{A} := L^{\infty-}(\Omega, \mu)$, where

$$L^{\infty-}(\Omega, \mu) := \bigcap_{1 \leq k < \infty} L^k(\Omega, \mu),$$

and define the state ω as

$$\omega(a) := \int_\Omega a(x) \mathrm{d}\mu(x), \quad a \in \mathcal{A}.$$

The tuple $(\mathcal{A}, \omega)$ defines a commutative probability space. Another standard example is the random matrices case. The algebra $\mathcal{A}$ consisting of $k \times k$ matrices over $L^{\infty-}(\Omega, \mu)$ is considered, where

$$\mathcal{A} := \mathcal{M}_k(L^{\infty-}(\Omega, \mu)).$$

The state ω on $\mathcal{A}$ is defined as

$$\omega(a) := \int_\Omega tr(a(x)) \mathrm{d}\mu(x), \quad a \in \mathcal{A},$$

where $tr(\cdot)$ is the normalized trace. The space $(\mathcal{A}, \omega)$ defines a non-commutative probability space, namely, the space of random matrices over (Ω, Σ, μ).

For a given non-commutative random variable $a \in \mathcal{A}$, the nth moments of a are provided by

$$m_n(\mu_a) := \int t^n \mathrm{d}\mu_a(t). \tag{A6}$$

Moreover, for given random variables $\{a_1, \cdots, a_n\}$ in $\mathcal{A}$, the moments are provided by

$$\omega(a_1 \cdots a_n) := \sum_{\pi \in \mathrm{NC}(n)} \kappa_\pi(a_1, \cdots, a_n), \tag{A7}$$

where κ_π are the *free cumulants*. The above equation is known as the *moments–cumulants formula*, where the free independence can be characterized by the vanishing of mixed cumulants (see [53], Chapter 11).

In free probability theory, a "convolution operation" can be defined for two free independent random variables $a_1, a_2 \in \mathcal{A}$. In most of this work, only the *free multiplicative convolution* is considered. Let a_1 and a_2 be two free independent random variables in $\mathcal{A}$ with their respective distributions μ_{a_1} and μ_{a_2}; it is assumed in what follows that μ_{a_1} and μ_{a_2} are compactly supported and that their support is a subset of $\mathbb{R}_+$. A free multiplicative convolution, or simply a *free product*, is defined by

$$\mu_{a_1 a_2} := \mu_{a_1} \boxtimes \mu_{a_2}, \tag{A8}$$

where $\mu_{a_1 a_2}$ represents the distribution of $a_1\, a_2$. As for the free additive convolution, there exists a standard and analytical way to compute the free product, which can be done via the *S-transform*. The *S-transform* of a probability distribution μ_a is defined as *the formal inverse of the moment-generating formal power series*, provided by

$$S_{\mu_a}(z) = \frac{1 - z}{z} M_a^{-1}(z), \tag{A9}$$

where $M_a^{-1}(z)$ is the formal inverse of the *moment-generating* formal power series provided by

$$M_{\mu_a}(z) := \sum_{k=1}^{\infty} m_{k,a}\, z^k. \tag{A10}$$

With the help of the S-transform, the free product can be computed, where:

$$S_{\mu_{a_1^{1/2}a_2a_1^{1/2}}}(z) = S_{\mu_{a_1}}(z)\, S_{\mu_{a_2}}(z). \tag{A11}$$

In the following, we recall some standard distributions that are well-known in the literature and highly used in this work.

The first distribution considered here is the *semicircular law*, which is one of the most important distributions encountered in free probability theory. The semicircular distribution $\mu_{\mathrm{SC}}(x)$ is defined by the density:

$$\mathrm{d}\mu_{\mathrm{SC}}(x) := \frac{\sqrt{4-x^2}}{2\pi}\mathbb{1}_{x\in[-2,2]}\mathrm{d}x. \tag{A12}$$

For illustration, the S-transform of the semicircular distribution can be computed by standard computation:

$$S_{\mu_{\mathrm{SC}}}(z) = \frac{-z+\sqrt{z^2-4}}{2}.$$

The first link that can be identified is that the moments of the semicircular distribution are intrinsically related to the Catalan numbers. The following equality holds:

$$\int x^k \mathrm{d}\mu_{\mathrm{SC}}(x) = \mathrm{Cat}_k \tag{A13}$$

where Cat_k are the Catalan numbers (see Equation (A1) and Chapter 2 in [53] for more details). Moreover, the S-transform provides another important link between free probability and combinatorics by allowing the computation of the moments of free product convolution of Marčhenko–Pastur distribution (see Theorem A5).

Another well-known result due to Wigner [56] shows the following link between the semicircular distribution and random Gaussian matrices.

Theorem A1 ([56])**.** *Let $N \in \mathbb{N}$ and let A_N be an $N \times N$ self-adjoint random Gaussian matrix. Then, A_N converges weakly to a semicircular distribution $\mu_{SC}(x)$.*

A complete proof can be found in [57]. In this particular case, a deep link is established between random Gaussian matrices, the semicircular law, and the Catalan numbers. Moreover, the semicircular law plays an important role in free probability theory as a free central limit distribution. Next, one of the main results in free probability theory is recalled (see Theorem 8.10 in [53]).

Theorem A2. *Let $(\mathcal{A}, \omega)$ be a non-commutative probability space and let $a_1, \cdots, a_N \in \mathcal{A}$ be free independent and identically distributed self-adjoint random variables. Assume that $\omega(a_i) = 0$ for $i \in [N]$ and denote the variance of the random variables a_i by $\sigma^2 := \omega(a_i^2)$. Then, the following holds:*

$$\frac{a_1 + \cdots + a_N}{\sqrt{N}} \to \mu_{SC}(x).$$

The sequence converges weakly to $\mu_{SC}(x)$ as $N \to \infty$, where s is a semicircular variable of variance σ^2.

This particular distribution helps to illustrate the relationship between random matrices, combinatorics, and free probability theory.

In the following, we provide another example of a distribution that plays an important role in this work. The second distribution considered here is the *Marčhenko–Pastur distribution*, denoted by $\mathrm{MP}(t)$ and defined by

$$\mathrm{MP}(t) := \max(1-t,0)\delta_0 + \nu_t,$$
$$d\nu_t(x) := \frac{\sqrt{4t-(x-1-t)^2}}{2\pi x}\mathbb{1}_{(x-1-t)^2\leq 4t}\mathrm{d}x.$$

The Marčhenko–Pastur distribution is deeply related to Wishart matrices. Let Z be a Wishart matrix defined as $Z := \frac{1}{m}YY^*$, where $Z \in \mathcal{M}_{nm}(\mathbb{C})$ and the entries of $Y \in \mathcal{M}_{nm}(\mathbb{C})$ are complex random Gaussian variables. It was shown by Marčhenko and Pastur that the empirical distribution of Wishart matrices converges to the $\mathrm{MP}(t)$ defined above. More precisely, the following theorem was established.

Theorem A3. *Consider a Wishart matrix Z and let $\mu_{n,m}$ be its empirical distribution provided by*

$$\mu_{n,m} := \frac{1}{n}\sum_{z\in\mathrm{spec}(Z)}\delta_z.$$

Assuming that n/m converges to t as $n \to \infty$, then $\mu_{n,m}$ converges (weakly) to $\mathrm{MP}(t)$ with $t > 0$.

For a proof and a more detailed statement of this result, see Theorem 3.6 and Theorem 3.7 in [58]. In particular, the case of MP is of importance in this work, where the distribution is as follows:

$$\mathrm{dMP} := \frac{1}{2\pi}\sqrt{4x^{-1}-1}\,\mathrm{d}x \tag{A14}$$

where the shorthand notation MP is used instead of $\mathrm{MP}(1)$.

For the semicircular distribution described previously, the moments of free convolution products of MP can be related to the Fuss–Catalan numbers. For conciseness, only some relevant results for the Marčhenko–Pastur distribution are stated here; for more details and proofs, see [48].

Theorem A4. *Let $\mathrm{MP}(t)$ be the Marčhenko–Pastur distribution. The S-transform is provided by*

$$S_{\mathrm{MP}(t)}(z) = \frac{1}{t+z}.$$

One of the main results of [48] was the relation between the free product convolution of $\mathrm{MP}(t)$ and combinatorics. Here, we state only the particular result for MP that is relevant for this work.

Theorem A5. *Let* MP *denote the Marčhenko–Pastur distribution. For $\mathrm{MP}^{\boxtimes s}$ with $s \geq 2$, the following holds:*

$$\int_{\mathbb{R}} x^n \mathrm{dMP}^{\boxtimes s} = \mathrm{FC}_{n,s} \tag{A15}$$

where $\mathrm{FC}_{n,s}$ are the Fuss–Catalan numbers.

References

1. Maldacena, J. The large-N limit of superconformal field theories and supergravity. *Int. J. Theor. Phys.* **1999**, *38*, 1113–1133. [CrossRef]
2. Ryu, S.; Takayanagi, T. Holographic derivation of entanglement entropy from the anti–de sitter space/conformal field theory correspondence. *Phys. Rev. Lett.* **2006**, *96*, 181602. [CrossRef]
3. Chen, B.; Czech, B.; Wang, Z.Z. Quantum information in holographic duality. *Rep. Prog. Phys.* **2022**, *85*, 046001. [CrossRef]
4. Cirac, J.I.; Perez-Garcia, D.; Schuch, N.; Verstraete, F. Matrix product states and projected entangled pair states: Concepts, symmetries, theorems. *Rev. Mod. Phys.* **2021**, *93*, 045003. [CrossRef]
5. Collins, B.; González-Guillén, C.E.; Pérez-García, D. Matrix Product States, Random Matrix Theory and the Principle of Maximum Entropy. *Commun. Math. Phys.* **2013**, *320*, 663–677. [CrossRef]
6. González-Guillén, C.E.; Junge, M.; Nechita, I. On the spectral gap of random quantum channels. *arXiv* **2018**, arXiv:1811.08847. [CrossRef]
7. Lancien, C.; Pérez-García, D. Correlation length in random MPS and PEPS. *Ann. Henri Poincaré* **2022**, *23*, 141–222. [CrossRef]
8. Hayden, P.; Nezami, S.; Qi, X.L.; Thomas, N.; Walter, M.; Yang, Z. Holographic duality from random tensor networks. *J. High Energy Phys.* **2016**, *2016*, 9. [CrossRef]
9. Dong, X.; Qi, X.L.; Walter, M. Holographic entanglement negativity and replica symmetry breaking. *J. High Energy Phys.* **2021**, *2021*, 24. [CrossRef]
10. Kudler-Flam, J.; Narovlansky, V.; Ryu, S. Negativity spectra in random tensor networks and holography. *J. High Energy Phys.* **2022**, *2022*, 76. [CrossRef]
11. Penington, G.; Shenker, S.H.; Stanford, D.; Yang, Z. Replica wormholes and the black hole interior. *J. High Energy Phys.* **2022**, *2022*, 205. [CrossRef]
12. Qi, X.L.; Shangnan, Z.; Yang, Z. Holevo information and ensemble theory of gravity. *J. High Energy Phys.* **2022**, *2022*, 56. [CrossRef]
13. Qi, X.L.; Yang, Z. Space-time random tensor networks and holographic duality. *arXiv* **2018**, arXiv:1801.05289. [CrossRef]
14. Qi, X.L.; Yang, Z.; You, Y.Z. Holographic coherent states from random tensor networks. *J. High Energy Phys.* **2017**, *2017*, 60. [CrossRef]
15. Yang, Z.; Hayden, P.; Qi, X.L. Bidirectional holographic codes and sub-AdS locality. *J. High Energy Phys.* **2016**, *2016*, 175. [CrossRef]
16. Levy, R.; Clark, B.K. Entanglement entropy transitions with random tensor networks. *arXiv* **2021**, arXiv:2108.02225. [CrossRef]
17. Lopez-Piqueres, J.; Ware, B.; Vasseur, R. Mean-field entanglement transitions in random tree tensor networks. *Phys. Rev. B* **2020**, *102*, 064202. [CrossRef]
18. Li, Y.; Vasseur, R.; Fisher, M.; Ludwig, A.W. Statistical mechanics model for Clifford random tensor networks and monitored quantum circuits. *arXiv* **2021**, arXiv:2110.02988. [CrossRef]
19. Medina, R.; Vasseur, R.; Serbyn, M. Entanglement transitions from restricted Boltzmann machines. *Phys. Rev. B* **2021**, *104*, 104205. [CrossRef]
20. Marolf, D.; Wang, S.; Wang, Z. Probing phase transitions of holographic entanglement entropy with fixed area states. *J. High Energy Phys.* **2020**, *2020*, 84. [CrossRef]
21. Nahum, A.; Roy, S.; Skinner, B.; Ruhman, J. Measurement and entanglement phase transitions in all-to-all quantum circuits, on quantum trees, and in Landau-Ginsburg theory. *PRX Quantum* **2021**, *2*, 010352. [CrossRef]
22. Vasseur, R.; Potter, A.C.; You, Y.Z.; Ludwig, A.W. Entanglement transitions from holographic random tensor networks. *Phys. Rev. B* **2019**, *100*, 134203. [CrossRef]
23. Yang, Z.C.; Li, Y.; Fisher, M.P.; Chen, X. Entanglement phase transitions in random stabilizer tensor networks. *Phys. Rev. B* **2022**, *105*, 104306. [CrossRef]
24. You, Y.Z.; Yang, Z.; Qi, X.L. Machine learning spatial geometry from entanglement features. *Phys. Rev. B* **2018**, *97*, 045153. [CrossRef]
25. Cheng, N.; Lancien, C.; Penington, G.; Walter, M.; Witteveen, F. Random Tensor Networks with Non-trivial Links. *Ann. Henri Poincaré* **2024**, *25*, 2107–2212. [CrossRef]
26. Apel, H.; Kohler, T.; Cubitt, T. Holographic duality between local Hamiltonians from random tensor networks. *J. High Energy Phys.* **2022**, *2022*, 52. [CrossRef]
27. Bao, N.; Penington, G.; Sorce, J.; Wall, A.C. Beyond toy models: Distilling tensor networks in full AdS/CFT. *J. High Energy Phys.* **2019**, *2019*, 69. [CrossRef]
28. Freedman, M.; Headrick, M. Bit threads and holographic entanglement. *Commun. Math. Phys.* **2017**, *352*, 407–438. [CrossRef]
29. Lewkowycz, A.; Maldacena, J. Generalized gravitational entropy. *J. High Energy Phys.* **2013**, *2013*, 90. [CrossRef]
30. Collins, B.; Nechita, I. Gaussianization and eigenvalue statistics for random quantum channels (III). *Ann. Appl. Probab.* **2011**, *21*, 1136–1179. [CrossRef]

31. Zhou, T.; Nahum, A. Emergent statistical mechanics of entanglement in random unitary circuits. *Phys. Rev. B* **2019**, *99*, 174205. [CrossRef]
32. De Luca, A.; Liu, C.; Nahum, A.; Zhou, T. Universality classes for purification in nonunitary quantum processes. *arXiv* **2023**, arXiv:2312.17744. [CrossRef]
33. Hastings, M.B. The asymptotics of quantum max-flow min-cut. *Commun. Math. Phys.* **2017**, *351*, 387–418. [CrossRef]
34. Harrow, A.W. The Church of the Symmetric Subspace. *arXiv* **2013**, arXiv:1308.6595. [CrossRef]
35. Collins, B.; Nechita, I. Random matrix techniques in quantum information theory. *J. Math. Phys.* **2016**, *57*, 015215. [CrossRef]
36. Korte, B.; Vygen, J. *Combinatorial Optimization*; Springer: Berlin/Heidelberg, Germany, 2011; Volume 1.
37. Billingsley, P. *Convergence of Probability Measures*; John Wiley & Sons: Hoboken, NJ, USA, 2013.
38. Bechet, D.; De Groote, P.; Retoré, C. A complete axiomatisation for the inclusion of series-parallel partial orders. In Proceedings of the International Conference on Rewriting Techniques and Applications, Barcelona, Spain, 2–5 June 1997; Springer: Berlin/Heidelberg, Germany, 1997; pp. 230–240.
39. Życzkowski, K.; Sommers, H.J. Induced measures in the space of mixed quantum states. *J. Phys. A Math. Gen.* **2001**, *34*, 7111. [CrossRef]
40. Sommers, H.J.; Życzkowski, K. Statistical properties of random density matrices. *J. Phys. A Math. Gen.* **2004**, *37*, 8457. [CrossRef]
41. Nechita, I. Asymptotics of random density matrices. *Ann. Henri Poincaré* **2007**, *8*, 1521–1538. [CrossRef]
42. Życzkowski, K.; Sommers, H.J. Hilbert-Schmidt volume of the set of mixed quantum states. *J. Phys. A Math. Gen.* **2003**, *36*, 10115. [CrossRef]
43. Życzkowski, K.; Penson, K.A.; Nechita, I.; Collins, B. Generating random density matrices. *J. Math. Phys.* **2011**, *52*, 062201. [CrossRef]
44. Page, D.N. Average entropy of a subsystem. *Phys. Rev. Lett.* **1993**, *71*, 1291. [CrossRef] [PubMed]
45. Foong, S.; Kanno, S. Proof of Page's conjecture on the average entropy of a subsystem. *Phys. Rev. Lett.* **1994**, *72*, 1148. [CrossRef] [PubMed]
46. Sánchez-Ruiz, J. Simple proof of Page's conjecture on the average entropy of a subsystem. *Phys. Rev. E* **1995**, *52*, 5653. [CrossRef]
47. Lancien, C. (Institut Fourier Grenoble, France). Personal communication, 2024.
48. Banica, T.; Belinschi, S.T.; Capitaine, M.; Collins, B. Free Bessel laws. *Can. J. Math.* **2011**, *63*, 3–37. [CrossRef]
49. Collins, B.; Nechita, I.; Życzkowski, K. Random graph states, maximal flow and Fuss–Catalan distributions. *J. Phys. A Math. Theor.* **2010**, *43*, 275303. [CrossRef]
50. Collins, B.; Nechita, I.; Życzkowski, K. Area law for random graph states. *J. Phys. A Math. Theor.* **2013**, *46*, 305302. [CrossRef]
51. Janson, S. *Gaussian Hilbert Spaces*; Cambridge University Press: Cambridge, UK, 1997.
52. Billingsley, P. *Probability and Measure*; Wiley Series in Probability and Statistics; Wiley: Hoboken, NJ, USA, 2012.
53. Nica, A.; Speicher, R. *Lectures on the Combinatorics of Free Probability*; Cambridge University Press: Cambridge, UK, 2006; Volume 13.
54. Mingo, J.A.; Speicher, R. *Free Probability and Random Matrices*; Springer: Berlin/Heidelberg, Germany, 2017; Volume 35.
55. Armstrong, D. Generalized Noncrossing Partitions and Combinatorics of Coxeter Groups. *arXiv* **2007**, arXiv:math/0611106. [CrossRef]
56. Wigner, E.P. Characteristic vectors of bordered matrices with infinite dimensions. In *The Collected Works of Eugene Paul Wigner: Part A: The Scientific Papers*; Springer: Berlin/Heidelberg, Germany, 1993; pp. 524–540.
57. Anderson, G.; Guionnet, A.; Zeitouni, O. *An Introduction to Random Matrices*; Cambridge Studies in Advanced Mathematics; Cambridge University Press: Cambridge, UK, 2010.
58. Bai, Z.; Silverstein, J.W. *Spectral Analysis of Large Dimensional Random Matrices*; Springer: Berlin/Heidelberg, Germany, 2010; Volume 20.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.