Published for SISSA by Springer

Received: *April 13, 2022*
Revised: *July 11, 2022*
Accepted: *September 2, 2022*
Published: *September 19, 2022*

Replicas, averaging and factorization in the IIB matrix model

Mihailo Čubrović

Center for the Study of Complex Systems,
Institute of Physics Belgrade, University of Belgrade,
Pregrevica 118, 11080 Belgrade, Serbia

E-mail: cubrovic@ipb.ac.rs

Abstract: We study the partition functions of multiple replicas (copies) of D-brane configurations in the type IIB (IKKT) matrix model. We consider the quenched regime, where small fluctuations of the matrices are superimposed onto the slow (quenched) dynamics of the background, so the partition function is an ensemble average over the background. Interacting D-branes always factorize in a simple way. On the other hand, the non-interacting BPS configurations may or may not factorize depending on the number of replicas, and their factorization mechanism is more involved as the corresponding saddle-point solutions (half-wormholes) break the replica symmetry. We argue that the simple factorization mechanism of interacting branes is actually more interesting as it carries the specific signatures of quantum gravity, which are absent from disordered field theories like the SYK model.

Keywords: AdS-CFT Correspondence, Matrix Models, Models of Quantum Gravity, Random Systems

ArXiv ePrint: 2203.10697

https://doi.org/10.1007/JHEP09(2022)136

JHEP09(2022)136

Contents

1 Introduction

The factorization problem in gravity and holography has come into the spotlight with the discovery of replica wormholes, entanglement islands and the path they open toward a possible solution of the black hole information paradox [1–3]. It also resonates with the general growth of knowledge in quantum information and its relation to black holes, scrambling [4–6] and quantum chaos [7, 8]. Stated simply, the puzzle lies in the fact that the spacetime (Euclidean)[1] wormholes constructed in [1, 2], which are crucial to save the Page curve and the unitarity of black hole evaporation, imply that the partition function

[1]Usually, the literature speaks of spacetime=Euclidean wormholes versus spatial=Lorentzian wormholes. But to be precise, Euclidean vs. Lorentzian signature is just a matter of choice, the true difference is between spatial wormholes which act as bridges of Kip Thorne style, and spacetime wormholes considered in this paper and in most works on factorization and averaging, which affect also the time direction. Therefore, even though the latter are usually considered in Euclidean time (also in this paper), one could also study them in Lorentzian signature. But one should still bear in mind that Euclidean/Lorentzian wormholes is the term often used in the literature, meaning really spacetime/ spatial wormholes.

of two copies (two replicas) of a gravitating system does not equal the square of a single-replica partition function: $Z_{1\mathrm{grav}}^2 \neq Z_{2\mathrm{grav}}$. But at least in asymptoticly anti-de-Sitter (AdS) geometries this contradicts the field theory intuition that the partition function of two identical decoupled systems should always factorize: $Z_{1\mathrm{CFT}}^2 = Z_{2\mathrm{CFT}}$. The way out proposed in several works [9–13] is that holography performs some kind of averaging or coarse-graning so that the dual CFT partition function is really an expectation value over some distribution:

$$\langle Z_{1\mathrm{CFT}} \rangle^2 \neq \langle Z_{2\mathrm{CFT}} \rangle. \tag{1.1}$$

In this case, the nonfactorization ceases being a puzzle — of course there is no reason that expectation values factorize. The averaging could in principle be carried over disorder ("explicit") or it could really be some kind of self-averaging in a chaotic system, i.e. some kind of coarse-graning.

A prototypical framework for explicit averaging is the Sachdev-Ye-Kitaev (SYK) model, featuring a system of Majorana fermions with all-to-all coupling, the coupling strengths being a quenched random variable. In a certain regime, this system is dual to gravity in AdS_2 [14], making it perfect for studying the factorization problem. A simplified version — SYK model in a single time point — was analyzed in [12, 15] and the outcome is very pleasing: although the wormhole configurations (which couple different replicas) are non-factorizing, there are additional solutions, dubbed half-wormholes, which restore the factorization; half-wormholes depend strongly on the choice of microscopic couplings. A similar picture was found to hold also in other systems like tensor models, random matrices and the two-site SYK model [16–19]. In [20–22] the traversable spatial ("Lorentzian") wormholes have also been related to field theories averaged over the states or over the operators, starting, as could be expected, from thermofield-dynamics (TFD)-like states. Some recent generalizations are found in [23–26] and in particular in [27], where the authors find that nonlocally interacting bulk branes in 2D gravity provide a mechanism which restores factorization in absence of any explicit disorder.

Our goal is to understand these workings of (half)wormholes in systems with an ensemble average but directly in quantum gravity, not in a field theory like SYK or a matrix model. The puzzle is now the following: is the averaging an operation which is somehow "automatically" performed specifically by holography, or we can perform it in a gravitating system directly, without any reference to the dual field theory (or for gravitating systems which have no dual CFT at all)? We will take a quantum gravity model (specifically the IIB string theory matrix model) and construct wormholes and half-wormholes by averaging over suitably chosen "quenched" degrees of freedom, thus repeating the logic of [9, 12, 15–17, 28, 29] but directly in gravity. Therefore, we mainly aim to understand factorization as such, not specifically in the context of AdS/CFT. But we will also discuss how one can make contact with AdS/CFT, by introducing a non-flat background metric; this is merely a first step and a full-fledged application of our findings to holography will be a subject of future work.

The arena for our work on wormholes and factorization is the model by Ishibashi, Kawai, Kitazawa and Tsuchiya (IKKT), proposed and developed in [30–33]. Like other

JHEP09(2022)136

matrix models of string and M theory (see [34, 35] for a review), it is potentially capable of providing a nonperturbative description of string theory, including D-branes and other deep quantum effects, which are beyond the scope of old-style perturbative string theory. On the other hand, as a matrix model, the IKKT system allows the explicit computation of observables and partition functions in a controlled way, and it has a lot in common with matrix models in field theory. In fact, in a certain limit the system we study can also be thought of as ten-dimensional Yang-Mills field theory with a quenched background field configuration [36, 37]. This makes our results relevant in principle also in the field theory context, and highlights that the formal workings of (non)factorization of partition functions are in a sense quite technical and independent of many physical details of the system.

Another important point is that the IKKT model is well-defined both in Lorentzian and Euclidean signature. While the latter is more convenient when studying the landscape of saddle-point solutions, as we can define a partition function in terms of the Euclidean action S_E in the usual way as $Z = \int \exp(-S_E)$, the former is more frequently studied in the literature, as the amplitude $\mathcal{A} = \int \exp(-\imath S_L)$ of the Lorentzian action S_L is always real; this is not the case for the partition function when the fermionic excitations are turned on. For example, the Lorentzian dynamics was argued to explain the effective 3+1-dimensionality of spacetime and other cosmologically relevant issues in [38–42]. We have opted for the Euclidean model so we can readily read off the free energy and determine which of the wormhole and half-wormhole solutions are thermodynamically prefered, but the reader should bear in mind that this is just a matter of convenience (see also the first footnote).

With some hindsight, we can say that a picture rather similar to the wormhole/half-wormhole story in SYK and similar models will emerge here. We have not found a single case where the factorization of $\langle Z^n \rangle$ is not restored already at leading order in perturbation theory for $n \geq 4$, although there are cases where the factorization is violated for $n = 2$. In some cases the factorization is trivial (when the averaged value $\langle Z^n \rangle$ is at leading order just the product of n copies of the leading-order estimate for $\langle Z \rangle$), and sometimes it is nontrivial, in the sense that it cannot be written in terms of $\langle Z \rangle$ contributions only. This distinction is interesting and depends on the geometry of the D-brane configuration. All of this is happening in (discretized) string theory, without any reference to holographic duality (although of course, one expects that the nonperturbative IKKT model implicitly knows about the duality).

As a final word of caution, one should bear in mind that more general wormhole configurations exist which are not necessarily all related to averaging, see, e.g. [44–47] and the recent insights of [48]. There is actually a very general geometric perspective on wormhole solutions, which works also in quantum mechanics (not only field theory), resting on the symplectic structure of the Hamiltonian dynamics [49]; this likewise suggests Euclidean wormholes to be more general than the averaging-induced configurations in this paper.

1.1 The sharp question

After all this talk, we are ready to formulate in a sharp way the main question of the paper — if the fluctuations of the D-brane solutions to the IKKT model factorize when averaged over the background fields. Factorization means that n replicas of the system behave in the same way as n independent copies. In precise language, this says that the averaged partition functions satisfy $\langle Z^n \rangle \sim \langle Z \rangle^n$. A related concept is self-averaging. Self-averaging means that the expectation value of $\langle Z^n \rangle$ is "close" to a "typical" Z^n value for some generic realization of the quenched variables (brane matrices). In precise language, denoting the quenched variables by λ, self-averaging is formulated as

$$\langle Z^n(\lambda) \rangle \sim Z^n \left(\sqrt{\langle \lambda^2 \rangle} \right). \tag{1.2}$$

The above means simply that the average over all λ values produces the same result at leading order as the value for an average λ; the reason we take the square root of $\langle \lambda^2 \rangle$ and not simply $\langle \lambda \rangle$ is that the latter is often zero (λ will often have a symmetric distribution).

The outline of the paper is the following. In section 2 we sum up the essential physics of D-branes in the type IIB matrix model and define the quenched approximation, setting the stage for the main work. Sections 3 and 4 contain the core of the paper: we find saddle-point solutions of Z^n, $\langle Z^n \rangle$ and $\langle Z \rangle^n$ and discuss their factorization properties, first for a simple, single-D-string configuration and then for interacting D-strings. Section 5 sums up the conclusions. In appendix A we show that a technical simplification made in the paper (putting the background fermions to zero) does not lead to any qualitative changes in our results. In appendix B we demonstrate another important technical point (that the conclusions do not depend qualitatively on whether the eigenvalue distribution is Gaussian or uniform). Finally, in appendix C we discuss how the reasoning of this paper could be applied specifically to the factorization puzzle of holography, in asymptotically AdS backgrounds.

2 Setup: D-brane configurations in the IKKT model

Let us start from the action of the type IIB matrix model, as found by Ishibashi, Kawai, Kitazawa and Tsuchiya [30, 31] by discretizing the Schild action for IIB superstrings:

$$S = -\mathrm{Tr} \left(\frac{1}{4} \left[X_\mu, X_\nu \right]^2 + \frac{1}{2} \bar{\Psi}_\alpha \Gamma^\mu \left[X_\mu, \Psi_\alpha \right] \right). \tag{2.1}$$

Here, $\mu = 1, \ldots 10$ are the spacetime (target space) dimensions and $\alpha = 1, \ldots 16$ counts the Majorana-Weyl fermions. The gamma matrices are then 16×16 matrices. Both the scalars and the spinors are $N \times N$ Hermitian matrices. The size N is also dynamic, corresponding to the auxiliary field g in the Schild action (see e.g. [34]). However, we will always work with fixed and large N, assuming as usual that the partition function is strongly dominated by a single saddle point at some N. For this reason there is no sum over N in the partition functions throughout the paper.

The equations of motion follow from (2.1):

$$[X^\mu, [X^\mu, X^\nu]] = 0, \quad [X^\mu, (\Gamma^\mu \Psi)_\alpha] = 0. \tag{2.2}$$

We work in the Euclidean signature, hence the spacetime metric is $\eta_{\mu\nu} = \mathrm{diag}(1, \ldots 1)$ and we do not need to differentiate between up and down indices. The partition function is now

JHEP09(2022)136

given in the usual way

$$Z = \sum_N \int D[X_\mu] \int D[\Psi_\alpha] \int D[\bar{\Psi}_\alpha] \exp(-S). \tag{2.3}$$

When only the bosonic degrees of freedom are excited, the above partition function is positive definite. If, however, the fermionic matrices Ψ are also nonzero, then their contribution to the path integral (the Pfaffian) is complex in Euclidean signature.[2] Therefore, we will need to be careful in interpreting the results when we turn on also the Ψ fields.

Now let us remember how D_p branes show up in the matrix model. What follows is a resume of the crucial aspects of brane solutions from [30, 34, 37]. Type IIB string theory admits D_p branes with p odd, starting from $p = -1$, i.e. D-instantons. Taking D-instantons as elementary degrees of freedom, we can write any configuration of size N as a superposition of N D-instantons at coordinates λ_j, $j = 1, \ldots N$. For a general D_p brane, the BPS condition and equations of motion from the action (2.1) lead to solutions $X_\mu = A_\mu$ for the bosonic part, with A_μ of the form:[3]

$$A_\mu = \left(\frac{L_1}{2\pi} q_1, \frac{L_2}{2\pi} k_1, \ldots \frac{L_{2i-1}}{2\pi} q_{(p+1)/2}, \frac{L_{2i}}{2\pi} k_{(p+1)/2}, 0 \ldots, 0 \right), \quad [q_i, k_i] \equiv \omega_i I = \imath \frac{L_{2i-1} L_{2i}}{2\pi N^{2/(p+1)}} I. \tag{2.4}$$

Here q_i and k_i are random Hermitian matrices with the eigenvalues $\lambda_j^{(q_i)}$ and $\lambda_j^{(k_i)}$ (of course, $p+1 \leq 10$), and L_μ ($\mu = 1 \ldots p+1$) are the compactification radii of the coordinates X_μ. The commutators ω_i have the meaning of $\hbar$, and $\omega_i \to 0$ corresponds to the classical limit — in this case q_i and k_i are just any commuting Hermitian matrices, describing the moduli of the theory. Multi-brane configurations are described by block-diagonal matrices q_i, k_i with M_p blocks of size $N \times N$ for M_p branes D_p. We will specialize to configurations of two D-strings as this suffices to discuss the factorization. The solution to the equations of motion (2.2) which corresponds to two strings at distance ℓ, with angle 2θ between the strings, reads [30]:

$$A_0 = \begin{pmatrix} q & 0 \\ 0 & q \end{pmatrix}, \quad A_1 = \begin{pmatrix} p\cos\theta & 0 \\ 0 & p\cos\theta \end{pmatrix}, \quad A_2 = \begin{pmatrix} \frac{\ell}{2} & 0 \\ 0 & -\frac{\ell}{2} \end{pmatrix}, \quad A_3 = \begin{pmatrix} -p\sin\theta & 0 \\ 0 & p\sin\theta \end{pmatrix}$$
$$A_4 = \ldots = A_{10} = 0, \quad [p, q] = \imath \frac{L^2}{2\pi N} I. \tag{2.5}$$

For $\theta = 0$ ($\theta = \pi$) we get two parallel (antiparallel) strings, and for $\ell = 0$ the strings cross. Parallel D-branes (including D-strings), with $\theta = 0$, are a special case: they are BPS states with zero on-shell action. This will be important as the properties and mechanisms of factorization differ between the BPS and the interacting case.

[2] This is not a problem by itself. Indeed, there lies the mechanism of the SO(10) symmetry breaking studied in [50, 51] in the context of the origin of the four-dimensionality of the spacetime.

[3] From now on, I always denotes the $N \times N$ unit matrix. If we need the unit matrix of a different size N', we will write it explicitly as $I_{N' \times N'}$.

2.1 The background-field regime and the quenched regime of the IKKT model

In the semiclassical regime one can study the fluctuations a_μ around the classical geometry described by matrices A_μ from (2.4) or (2.5):

$$X_\mu = A_\mu + a_\mu, \quad \Psi_\alpha = \psi_\alpha, \tag{2.6}$$

where we have taken into account that the background configuration of the fermions is $\Psi_\alpha = 0$. This spoils the supersymmetry and non-renormalization results, but does not change the essence of our conclusions; indeed, in the large N limit this is physically justified as the fermions are large-N suppressed anyway.[4] In appendix A we redo the main calculation of the paper in the presence of nonzero background Ψ_α and show that the outcome is essentially the same, though different in quantitative details of the solution. In the classical (commuting) regime, the matrices A_μ can be diagonalized simultaneously, their eigenvalues having the role of locations of D-instantons. In applications to D-brane dynamics and other nonperturbative string theory phenomena, as well as gravity and the dynamics of spacetime dimensions, the natural viewpoint is the background field approach, i.e. to integrate out the quantum fluctuations a_μ, ψ_α and study the effective semiclassical action for the diagonal matrices A_μ.

The opposite limit is the quenched IKKT model of [37], where the moduli A_μ are considered as fixed backgrounds and not integrated over. Now a_μ (and ψ_α, if turned on) are "dynamical" variables, and A_μ are quenched, so the partition function Z is tied to a specific realization of A_μ. This approach is natural if we want to study the quantum dynamics of perturbative string excitations in a given geometry. It can also be interpreted in the context of the Eguchi-Kawai finding [36] that the Lagrangian of the Yang-Mills gauge theory in the large N limit loses the derivative terms and reduces just to (2.1). In that context the dynamics of a_μ and ψ_α is basically the dynamics of a Yang-Mills system in the presence of quenched disorder.[5] In other words, *the fields A_μ play the same role as disorder in the SYK and similar models, describing on one hand a prototypical strongly interacting large N field theory and on the other hand string theory in deep nonperturbative regime.* This is our motivation to study the quenched IKKT model. Furthermore, AdS/CFT suggests that the background-field limit and the quenched limit show different aspects of the same physics — the Eguchi-Kawai system must have a dual description in terms of D-brane stacks. Therefore, the quenched IKKT model is a robust system of equations with several important physical aspects and should provide us an excellent testing ground for the factorization puzzle. From now on, we will denote the non-averaged quantities by ordinary letters, e.g. Z for the partition function, and the quantities averaged over the realizations

[4] Another way to say this is that the fermionic background can be put to zero in the semiclassical limit. Our general formalism is not limited to the semiclassical regime, but the physical picture does become simpler in that case, as we can interpret the matrix entries as spacetime coordinates of the constituent D-instantons making the D-brane. Such approximation is also exploited for similar reasons in some of the original IKKT papers [30, 32].

[5] The equivalence is not complete, as the quenched IKKT model contains zero modes, from the terms which couple a^μ and ψ_α; such terms do not exist in the large-N Yang-Mills theory. However, at leading — quadratic — order such terms vanish also in the quenched IKKT model (see eqs. (2.8)–(2.9)).

of the quenched variables A_μ will be denoted as expectation values, inside angular brackets, e.g. $\langle Z\rangle$.

Let us now write the IKKT action (2.1) in terms of the background field A_μ and the fluctuations a_μ, ψ_α, as in (2.6). The quadratic (S_2) and quartic (S_4) contributions in a_μ and ψ_α read:

$$S = S_2 + S_4 \tag{2.7}$$

$$S_2 = \mathrm{Tr}\left[-\frac{1}{4}a_\mu\left(P^2\delta_{\mu\nu}+2F_{\mu\nu}\right)a_\nu + \bar{\psi}_\alpha \not{P}\psi_\alpha - \bar{c}P^2 c\right], \tag{2.8}$$

$$S_4 = \mathrm{Tr}\left[\bar{\psi}_\alpha \not{a}\psi_\alpha - 2\left(P_\mu a^\mu\right)\left(a^\nu a_\nu\right) - \frac{1}{2}\left[a^\mu, a^\nu\right]\left[a_\mu, a_\nu\right]\right], \tag{2.9}$$

where we have introduced the superoperators P_μ and $F_{\mu\nu}$, acting on matrices as

$$P_\mu \equiv \left[A_\mu, \cdot\right], \quad F_{\mu\nu} = \left[\left[A_\mu, A_\nu\right], \cdot\right], \quad P^2 = P_\mu P^\mu. \tag{2.10}$$

We will also sometimes use the matrix $f_{\mu\nu} \equiv [A_\mu, A_\nu]$. Finally, c in (2.8) is the ghost field arising from the Faddeev-Popov ghost action. BPS states have $f_{\mu\nu} = \mathrm{const.}\times I$ with I being the unit matrix or, equivalently, $F_{\mu\nu} = 0$; in that case the on-shell action vanishes, one-loop quantum corrections are absent and the eigenvalues of A_μ remain uniformly distributed.

As we explained, for our purposes it is enough to focus on the bosonic sector and put the background fermionic matrices to zero: $\Psi_\alpha = 0$. This is always a consistent (if not fully generic) solution of the equations of motion, and it simplifies the calculations significantly. Bose-only backgrounds always give real action but they are obviously not protected by supersymmetry so even if $F_{\mu\nu} = 0$ there will be a logarithmic attraction of the eigenvalues as there is no fermionic sector to cancel the determinant from the bosonic path integral; however it remains (trivially) true that the solutions with $F_{\mu\nu} = 0$ describe a non-interacting configuration (in the Eguchi-Kawai picture, the field strength is zero). In general, even in zero fermionic background there are still nonzero fermionic fluctuations ψ_α. As a special case, we can turn off these too: $\psi_\alpha = 0$. In that case only the first term in (2.8) is nonzero, and in (2.9) the second and the third term remain nonzero. We will consider this special case as a warmup but will always include also the fermionic fluctuations ψ_α in the end.

3 Replicas and factorization for a single D-string

Now that we have set up the formal framework, we can explore the main question of the paper: how do the partition functions of D-branes in the quenched IKKT model factorize? From now on, we will specialize to D-strings as the whole story remains the same for higher-dimensional branes. In this section we give a detailed calculation for a single bosonic D-string, so the reader can get acquainted with the basic algorithm on the simplest example. Afterwards it will be straightforward to redo the calculation for the more interesting configurations of interacting strings.

3.1 Single copy — direct calculation

As we know [34], the background fields A_μ representing a D-brane are described by random Hermitian matrices p, q which, according to (2.4), satisfy the commutation relation (putting

$L_1 = L_2 \equiv L$):

$$[p, q] = \omega = \frac{L^2}{2\pi N} I. \tag{3.1}$$

This is a BPS state (although we do not see the supersymmetry when we put $\Psi_\alpha = 0$) with $f_{21} = -f_{12} = \omega$ and $F_{\mu\nu} = 0$. The eigenvalues λ^i_μ ($\mu = 1, 2$, $i = 1 \dots N$) are distributed in the interval $-L \leq \lambda_i \leq L$. Unlike the typical choice in the IKKT model literature, where the compactification radii L_μ define the hard cutoffs of the distribution, we adopt — solely for computational reasons — a soft cutoff with the Gaussian distribution of eigenvalues (Gaussian Unitary Ensemble) for A_μ so the width of the Gaussian equals L_μ. Our choice is unusual but on one hand it is no less physical (indeed, a hard cutoff is more of an idealization than a continuous Gaussian tail) and on the other hand more convenient for calculations. However, we emphasize that nothing changes qualitatively even when we return to hard cutoffs — in appendix B we show this explicitly, redoing the calculations with hard cutoff regularization.

From (2.3) we now have the following form for the partition function in the non-averaged and in the averaged form, respectively:

$$Z = \int D[a_\mu] e^{-S(a_\mu; A_\mu)} \tag{3.2}$$

$$\langle Z \rangle = \int D[A_\mu] \int D[a_\mu] e^{-S(a_\mu; A_\mu)} \mathcal{P}(A_\mu) = \int D[a_\mu] \int d^{2N} \lambda^i_\mu e^{-S\left(a_\mu; \lambda^i_\mu\right) - \frac{1}{2L^2} \lambda^i_\mu \lambda^{\mu i}}. \tag{3.3}$$

In the second equality, we have expressed A_μ in terms of its eigenvalues λ^i_μ and likewise the Gaussian measure (regulator) $\mathcal{P}$ is written out explicitly in terms of the width of the distribution L, as $-\lambda_\mu \lambda^\mu / 2L^2$. Now we want to write out the action in (3.3) in terms of matrices $A_0 = p$ and $A_1 = q$. It is easiest to diagonalize the matrices and work in the eigenbases. Then we have $A_\mu = \mathrm{diag}(\lambda^\mu_1, \dots \lambda^\mu_N)$. The superoperators P_μ are now represented as[6]

$$\begin{aligned} P_1 &= q \otimes I - I \otimes q, \quad (P_1)^{ij}_{kl} = q_{ik}\delta_{jl} - q_{jl}\delta_{ik} \\ P_2 &= k \otimes I - I \otimes k, \quad (P_2)^{ij}_{kl} = k_{ik}\delta_{jl} - k_{jl}\delta_{ik}\,. \end{aligned} \tag{3.4}$$

However, since the canonical momentum and the conjugate coordinate do not commute, we cannot assume both matrices to be diagonal at the same time, i.e. in the same basis. The above representation (3.4) takes each P_μ in its own eigenbasis. In order to give the reader a clearer intuitive grasp, we can list a few supermatrix elements, say for P_1:

$$(P_1)_{11} = \begin{pmatrix} 0 & p_{12} & \dots & p_{1N} \\ p_{21} & p_{22} - p_{11} & \dots & p_{2N} \\ \dots & \dots & \dots & \dots \\ p_{N1} & p_{N2} & \dots & p_{NN} - p_{11} \end{pmatrix}$$

$$(P_1)_{12} = \mathrm{diag}(-p_{12}, \dots - p_{12}), \quad (P_1)_{1N} = \mathrm{diag}(-p_{1N}, \dots - p_{1N})$$

[6]Superoperators act on $N \times N$ matrices, hence they are strictly speaking the Kronecker products of two $N \times N$ matrices. For convenience, we write them simply as $N^2 \times N^2$ matrices.

JHEP09(2022)136

$$\dots$$

$$(P_1)_{21} = \mathrm{diag}(-p_{21}, \dots - p_{21}), \;\; (P_1)_{22} = \begin{pmatrix} p_{11} - p_{22} & p_{12} & \dots & p_{1N} \\ p_{21} & 0 & \dots & p_{2N} \\ \dots & \dots & \dots & \dots \\ p_{N1} & p_{N2} & \dots & p_{NN} - p_{22} \end{pmatrix}$$

$$(P_1)_{N,N-1} = \mathrm{diag}(-p_{N,N-1}, \dots - p_{N,N-1})$$

$$(P_1)_{NN} = \begin{pmatrix} p_{11} - p_{NN} & p_{12} & \dots & p_{1N} \\ p_{21} & p_{22} - p_{NN} & \dots & p_{2N} \\ \dots & \dots & \dots & \dots \\ p_{N1} & \dots & p_{N,N-1} - p_{NN} & 0 \end{pmatrix}. \tag{3.5}$$

Averaging over the quenched variables now requires either adopting a single basis and transforming all P_μ but one (say P_1) from the form (3.4) to this fixed basis, or keeping each P_μ in its own eigenbasis — but then we have to divide the integral measure by the volume of the unitary matrices that transform each supermatrix to its eigenbasis. The latter is more convenient, and it gives rise to the eigenvalue attraction term. Plugging in the representation (3.4) into Z from (3.3) and taking into account the basis change we get

$$\begin{aligned} \langle Z \rangle &= \int D[a_\mu] \int d^{2N} \lambda_{\mu i} \Pi_{i<j} \left(\lambda_{\mu i} - \lambda_{\mu j}\right)^2 \exp\left[-\frac{1}{4} a^\dagger_{\mu i j} \left(\lambda^2_{\mu i} + \lambda^2_{\mu j}\right) a_{\mu k l} \delta_{jk} \delta_{il} - \frac{1}{2L^2} \lambda^2_{\mu i}\right] \\ &= \int D[a_\mu] e^{-W_1}, \quad W_1 = \frac{1}{2} \sum_\mu \log\det\left(\frac{I}{L^2_\mu} + 2a^\dagger_\mu a_\mu - 2I \mathrm{Tr} a^\dagger_\mu a_\mu\right). \end{aligned} \tag{3.6}$$

We have introduced the effective macroscopic action W_1 akin to the thermodynamic free energy. The first line in (3.6) is obtained by inserting the matrix representations for P into the general expression (3.3) taking into account the basic change (see e.g. [32]), and the final expression for W_1 in the second line comes from performing the Gaussian integral over λ^μ_i and the second-order expansion of the determinant in small $a^\dagger a$ and $1/L^{2N}$. Since the partition function is not a power-law function of a and $a^\dagger$, the series expansion has correlation functions $a^\dagger a \dots a^\dagger a$ of arbitrarily high order. Their meaning is better grasped in the collective field formalism that we introduce in the next subsection.

3.2 Collective fields and replicas

3.2.1 Warmup: single partition function again

The time is ripe to introduce the collective fields. We follow the formalism of [12, 15, 17] and largely adopt the notation of [12]. The idea is the following: we define a bilinear operator g as being equal to the current $a^\dagger a$. We impose this equality as a constraint through the Dirac delta functional, and finally replace the $a^\dagger a$-dependent terms by the appropriate functionals of the bilinear:[7]

$$\begin{aligned}\langle Z\rangle &= \int D[a]\int D[g]\int d\lambda \exp\left[-\frac{1}{2}\mathrm{Tr}a^\dagger P^2 a - \frac{2\left(\mathrm{Tr}g - \mathrm{Tr}a^\dagger a\right)}{L^{2N-2}}\right]\delta\left(g - a^\dagger a\right)\mathcal{P}(\lambda)\\ &= \int D[a]\int D[g]\int D[s]\int d\lambda \exp\left[-\frac{1}{2}\mathrm{Tr}a^\dagger P^2 a - \frac{2\left(\mathrm{Tr}g - \mathrm{Tr}a^\dagger a\right)}{L^{2N-2}} - \imath \mathrm{Tr}\left[s^\dagger\left(g - a^\dagger a\right)\right]\right]\mathcal{P}(\lambda)\\ &= \int D[a]\int D[g]\int \frac{D[s]}{2^N \pi^{N^2}}\exp\left[-\frac{1}{2}\log\det\left(\frac{I}{L^2} + 2a^\dagger a - 2I\mathrm{Tr}a^\dagger a\right) - \frac{2\left(\mathrm{Tr}g - \mathrm{Tr}a^\dagger a\right)}{L^{2N-2}}\right.\\ &\quad \left. - \imath \mathrm{Tr}\left[s^\dagger\left(g - a^\dagger a\right)\right]\right] \sim \int D[g]\int \frac{D[s]}{2^N \pi^{N^2}}\exp\left[-\frac{1}{2}\log\det s - \imath \mathrm{Tr}s^\dagger g - \frac{2}{L^{2N-2}}\mathrm{Tr}g\right]. \end{aligned} \tag{3.7}$$

In the first line, we have inserted into the effective action the term proportional to $\mathrm{Tr}g - \mathrm{Tr}a^\dagger a$ which, at leading order in $a^\dagger a$, equals zero on-shell, i.e. when the constraint $g = a^\dagger a$ is obeyed. In the second line, we have implemented the Dirac delta in the usual way, through the auxiliary field s, and in the third line we have averaged over the eigenvalues of the quenched coordinates λ. Notice how the inserted "zero term" precisely cancels all $a^\dagger, a$ dependence after averaging, which was of course the reason to introduce it in the first place. Again, this only holds at leading order in the a-fields and $1/L^N$. This is important — all our calculations are perturbative (unlike the simpler SYK quantum mechanics where the collective fields can be introduced in an exact way); for example, the last line in (3.7) holds at the order $O\left(\left(a^\dagger a\right)^4\right) + O(1/L^{4N})$. We will not write explicitly such higher-order remainders. Higher order terms could be taken into account perturbatively, by introducing additional collective fields and eventually closing the series by expressing the $l+1$-st order terms in terms of the loop integrals over the l-th and lower order terms. We will do this explicitly for the $\langle Z^2\rangle$ calculation, when it will be necessary to get anything nontrivial. For now, let us stay at the lowest order.

Now we can look for the saddle-point solutions of the effective action W_1 defined by the last line in (3.7) in the same way as in (3.6), through $W_1 \equiv -\log\langle Z\rangle$:

$$\begin{aligned} \frac{\partial W_1}{\partial g} &= \frac{2}{L^{2N-2}}I + \imath s = 0, & \frac{\partial W_1}{\partial s} &= \imath g + \frac{1}{2}\left(s^{-1}\right)^T = 0\\ s &= \frac{2\imath}{L^{2N-2}}I, & g &= \frac{L^{2N-2}}{4}I. \end{aligned} \tag{3.8}$$

Here, I is the $N\times N$ unit matrix as usual. The solution (3.8) is unique.

The important conclusion is that the nonzero solution (3.8) consists of scalar matrices which scale as L^{2N-2}, and inserting them into W_1 yields

$$W_1|_{\mathrm{on-shell}} \sim (N^2 - N)\log L + \frac{N}{2}\log 2. \tag{3.9}$$

Here and in the future we routinely disregard the contributions which go to zero when $N\to\infty$ or $L\to\infty$; we will only write them in a few special occasions when we want

[7] For the sake of brevity we leave out the spatial indices μ, ν etc. from now on. We will only write them when leaving them out could lead to confusion.

to emphasize that some part of the partition function contributes only negligible terms. One can check that the same value follows from the expression (3.6) after we solve the saddle-point equation for a_μ. We will compare this value with the on-shell value of the effective action for $\langle Z^2\rangle$ to check for factorization. One final remark: there is always a sum over the spacetime coordinate μ (in general $\mu = 1,\dots 2p$, in our case $\mu = 1,2$), hence W_1 from (3.9) is really multipled by 2 ($2p$ for a general D_p brane). But we do not write this factor explicitly in order not to clutter the notation; our W_1 (and similar for the two- and four-replica actions W_2 and W_4 we are yet to compute) is really the effective action per spacetime dimension (and since we work in Euclidean signature all dimensions are equivalent).

3.2.2 Double and quadruple partition function

Now that we have the averaged partition function $\langle Z\rangle$ for a D-brane, our question is: does the replicated partition function $\langle Z^2\rangle$ factorize? With two copies (replicas), the collective fields show their true meaning. The replicated partition function is[8]

$$Z^2 = \int D[a] \exp\left[-\mathrm{Tr}\left(2\lambda^i\left(a_A^\dagger a_A\right)\lambda^i + \frac{I}{2L^2}\lambda_i^2\right)\right], \tag{3.10}$$

where now we have the left and right replicas, denoted by the indices $A, B, \dots$ taking values L or R. The fun part is the averaged function $\langle Z^2\rangle$:

$$\begin{aligned}\langle Z^2\rangle &= \int D[a]\int D[g]\int\frac{D[s]}{2\pi}\int d\lambda e^{-\frac{1}{2}\mathrm{Tr}a_A^\dagger P^2 a_A} e^{-V(g_{AA})+V\left(a_A^\dagger a_A\right)} e^{-\imath\mathrm{Tr}\left[s_{AA}^\dagger\left(g_{AA}-a_A^\dagger a_A\right)\right]}\mathcal{P}(\lambda)\\ &= \int D[g]\int\frac{D[s]}{2^N\pi^{N^2}}\exp\left[-\imath\mathrm{Tr}\left(s_{AA}^\dagger g_{AA}\right)-\frac{1}{2}\log\det s_{AA}-V\left(g_{AA}\right)\right]\equiv e^{-W_2}.\end{aligned} \tag{3.11}$$

Analogously to what we did for $\langle Z\rangle$, the term with $-V(g_{AA}) + V\left(a_A^\dagger a_A\right)$ in the exponent is the multiplication by unity on-shell, and the s_{AA} auxiliary fields implement the Dirac delta functional. Up to fourth order, the interaction terms V read (from performing the Gaussian integration over λ_i and expanding the determinant):

$$\begin{aligned} V &= V_2 + V_4\\ V_2 &= \frac{2}{L^{2N-2}}\mathrm{Tr}(g_{LL}+g_{RR})\\ V_4 &= \frac{4}{L^{2N-4}}\mathrm{Tr}^2(g_{LL}+g_{RR}) - \frac{4}{L^{2N-4}}\mathrm{Tr}\left(g_{LL}^2 + g_{LL}g_{RR} + g_{RR}g_{LL} + g_{RR}^2\right). \end{aligned} \tag{3.12}$$

Again, this is just the expansion of the averaged function $\langle Z^2\rangle$ to fourth order in a_A, i.e. to second order in g_{AA}; the full expansion of the $\log\det$ term is infinite. This is a consequence of the initial action S_{eff} being quadratic in the fields, so the effective actions W_n contain the $\log\det$ term. If we had a linear coupling to the sources of the form Ja, we would get the

[8] In order to make the equations more compact we do not write out the replica indices in the measure of the integrals. Therefore, we write $\int D[g]$ for $\int D[g_{AA}]$, or $\int D[a]$ for $\int D[a_A]$ and similar. We still write the replica indices in the integrands (unlike the spatial indices $\mu, \nu \dots$ which we usually leave out.)

JHEP09(2022)136

by now familiar wormholes with collective fields g_{LR}, as in the SYK model in [11, 15, 17]. Now the saddle point equations from (3.12) are:

$$\frac{1}{2}\left(s_{AA}^{-1}\right)^{T}+\imath g_{AA}=0\ \Rightarrow\ s_{AA}=\frac{\imath}{2}\left(g_{AA}^{T}\right)^{-1} \tag{3.13}$$

$$\imath s_{AA}+\frac{2I}{L^{2N-2}}+\frac{8I}{L^{2N-4}}\mathrm{Tr}(g_{LL}+g_{RR})-\frac{8}{L^{2N-4}}(g_{LL}+g_{RR})^{T}=0. \tag{3.14}$$

While complicated at first glance, this system of equations has a high degree of symmetry. First, it is manifestly $L\leftrightarrow R$ invariant; second, it allows (though does not require) the maximally symmetric ansatz, where s and g are scalar matrices. Their matrix indices are $i,j\in\{1,\dots N\}$ (which we do not write explicitly) and $A,B\in\{L,R\}$. Therefore a scalar solution assumes (1) the $\mathrm{U}(N)$ group of the matrix model is fully preserved (which is logical in the large-N, random matrix regime that we study) and (2) full *replica symmetry* is preserved. The latter can be broken, and later on we will discuss this possibility in detail; but for now let us look at the replica-symmetric solution. We first insert s from (3.13) into (3.14) and then take the trace of both sides. Denoting $\mathrm{Tr}g_{LL}=\mathrm{Tr}g_{RR}\equiv t$, we find

$$t\sim-\frac{1}{16L^2}\pm\frac{L^{N-2}}{4\sqrt{2}\sqrt{N}}. \tag{3.15}$$

At leading order in large N and L, the two traces are symmetric, behaving as $\mathrm{Tr}g_{LL}=\mathrm{Tr}g_{RR}\sim\pm L^{N-2}/\sqrt{N}$. Still assuming maximal symmetry, this gives the following solution:

$$s=\pm\frac{2\imath}{t}I\otimes E,\quad g=\pm tI\otimes E. \tag{3.16}$$

Here, E is the two-by-two unit matrix in the replica space; its indices are $A,B\in\{L,R\}$. From now on we use this Kronecker product notation for the collective fields with replica indices: a field g_{AA} can be written in the form $B\otimes C$, with B being an $N\times N$ matrix and C being an $n\times n$ matrix with n the number of replicas.

The solutions scale as L^{N-2} instead of L^{2N-2} for the single replica solution (3.8). This fact already suggests that the two-replica solution does not factorize. Inserting (3.16) into (3.12) we find (for either sign in (3.16)):

$$W_2|_{\mathrm{on-shell}}\sim 2N^2\log L+L^{2N}/2\pm(2N)^{3/2}L^{N+2}. \tag{3.17}$$

Comparing to (3.9), we see that indeed *the two-replica solution does not factorize:* $\langle Z^2\rangle\neq\langle Z\rangle^2$. The first term in (3.17) precisely equals $2W_1$,[9] but it is actually strongly subleading compared to the other two terms, which behave entirely differently, as L^N and L^{2N}. Therefore, we can spot a contribution equaling $2W_1$ but there is also a much larger contribution absent in W_1.

So the partition function $\langle Z^2\rangle$ is not factorizing. Is it self-averaging? To check this, we can again follow [12] and compute the four-replica solution. Exploiting the definition of the

[9]Remember that all the solutions and on-shell action values are calculated for large N and L, meaning we disregard the terms which tend to zero as $N\to\infty$. Such higher-order corrections would likely spoil the exact ratio but this is expected.

non-averaged Z^2 from (3.11) we write:

$$
\begin{aligned}
Z^2 &= \int \frac{D[s]}{2^N \pi^{N^2}} \chi(s;g)\Phi(s) \\
\chi(s;g) &= \int D[g] \exp\left[-\imath \mathrm{Tr}\left(s^\dagger_{AA} g_{AA}\right) - V\left(g_{AA}\right)\right] \\
\Phi(s) &= \int D[a] \exp\left[\imath \mathrm{Tr}\left(s^\dagger_{AA} a^\dagger_A a_A\right) - \frac{1}{2}\mathrm{Tr} a^\dagger_A P^2 a_A + V\left(a^\dagger_A a_A\right)\right].
\end{aligned} \tag{3.18}
$$

Here we have separated the partition function into the non-averaging part $\chi(s;g)$ which just doubles when computing $\langle Z^4 \rangle$ as it is independent of a_A, and the part $\Phi(s)$ with a_A-dependent integrand which gets nontrivial additional correlations when Z^4 is averaged. Now the replica indices A, B can take values L, R, and A', B' take values L', R'. In addition to the two-replica fields, we now also have combinations of the form $LLR'R'$ and similar:[10]

$$
\begin{aligned}
\langle \Phi^2(s) \rangle &= \int D[a_A] \int D[a_{A'}] \int d^N \lambda \mathcal{P}(\lambda_i) \\
&\quad \times \exp\left[\imath \mathrm{Tr}\left(s^\dagger_{AA} a^\dagger_A a_A + s^\dagger_{A'A'} a^\dagger_{A'} a_{A'} - \frac{1}{2} a^\dagger_A P^2 a_A - \frac{1}{2} a^\dagger_{A'} P^2 a_{A'}\right) + V\left(a^\dagger_A a_A\right) + V\left(a^\dagger_{A'} a_{A'}\right)\right] \\
&= \int D[a_A] \int D[a_{A'}] \exp\left[\imath \mathrm{Tr}\left(s^\dagger_{AA} a^\dagger_A a_A + s^\dagger_{A'A'} a^\dagger_{A'} a_{A'}\right) + V\left(a^\dagger_A a_A\right) + V\left(a^\dagger_{A'} a_{A'}\right)\right] \\
&\quad \times \exp\left[-\frac{1}{2} \log\det\left(\frac{I}{L^2} + 2 a^\dagger_A a_A + 2 a^\dagger_{A'} a_{A'} - 2I \mathrm{Tr} a^\dagger_A a_A - 2I \mathrm{Tr} a^\dagger_{A'} a_{A'}\right)\right].
\end{aligned} \tag{3.19}
$$

The next step is to expand the argument of the logarithm, which results in the mixing of L, R and L', R'. Expanding to order four, we generate a new interaction potential, mixing all four replicas, and again resort to the trick of multiplying by unity and introducing the four-replica collective field $G_{AAB'B'}$, together with the Dirac delta constraint implemented by $S_{AA'BB'}$:

$$
\begin{aligned}
\langle \Phi^2(s) \rangle &= \int D[a_A] \int D[a_{A'}] \int D[S] \int D[G] \\
&\quad \times \exp \mathrm{Tr}\left[\imath s^\dagger_{AA} a^\dagger_A a_A + \imath s^\dagger_{A'A'} a^\dagger_{A'} a_{A'} + \imath S^\dagger_{AAB'B'}\left(a^\dagger_A a_A a^\dagger_{B'} a_{B'} - G_{AAB'B'}\right)\right] \\
&\quad \times \exp\left[V\left(a^\dagger_A a_A\right) + V\left(a^\dagger_{A'} a_{A'}\right) - \mathcal{V}\left(g_{AA}, g_{B'B'}; G_{AAB'B'}\right) + \mathcal{V}\left(a^\dagger_A a_A, a^\dagger_{B'} a_{B'}\right)\right] \\
&= \int D[S] \int D[G] e^{-\tilde{W}_4},
\end{aligned} \tag{3.20}
$$

with the effective action that now depends also on the new collective fields $G_{AAB'B'} \equiv a^\dagger_A a_A a^\dagger_{B'} a_{B'}$. From (3.18), the total effective action is the sum of $\tilde{W}_4$ from the last equation and the non-averaging contribution from $\chi(s;g)$:

$$
\begin{aligned}
W_4 &= \imath \mathrm{Tr}\left(s^\dagger_{AA} g_{AA}\right) + V\left(g_{AA}\right) + \tilde{W}_4 \\
\tilde{W}_4 &= \frac{1}{2}\left(\log\det s_{AB} + \log\det s_{A'B'} + \log\det S_{AAB'B'}\right) - \imath \mathrm{Tr}\left(S^\dagger_{AAB'B'} G_{AAB'B'}\right) + \mathcal{V} \\
\mathcal{V}\left(g_{AA}, g_{B'B'}; G_{AAB'B'}\right) &= \frac{8}{L^{2N-4}} \mathrm{Tr} g_{AA} \mathrm{Tr} g_{B'B'} - \frac{4}{L^{2N-4}} \mathrm{Tr} G_{AAB'B'}.
\end{aligned} \tag{3.21}
$$

[10] In this equation, exceptionally, we write the indices in the integrand, e.g., $\int D[a_{A'}]$, in order to emphasize that we now have extra copies of the fields.

This is the main formal result of this section. We will now consider various saddle-point solutions of the effective action (3.21), and consider which of these may restore the factorization and which are self-averaging.

3.3 Half-wormholes

The effective action (3.21) is slightly more involved but the highest symetry solutions are still easy to find. The saddle-point equations read:

$$
\begin{aligned}
&-\frac{1}{2}\left(g_{AA}^{-1}\right)^T+\frac{2I}{L^{2N-2}}+\frac{8I}{L^{2N-4}}\mathrm{Tr}(g_{LL}+g_{RR})-\frac{8}{L^{2N-4}}(g_{LL}+g_{RR})^T-\frac{8I}{L^{2N-4}}\mathrm{Tr}g_{B'B'}=0\\
&\frac{1}{2}\left(S_{AAB'B'}^{-1}\right)^T+\imath G_{AAB'B'}=0,\quad -\imath S_{AAB'B'}-\frac{4}{L^{2N-4}}I=0.
\end{aligned}\tag{3.22}
$$

Notice that the first and the second line are decoupled. The second line (the equations for S and G) is linear and the solution is unique and left-right symmetric in the replica space (invariant to $L\leftrightarrow R$ and $L'\leftrightarrow R'$). In the first line, we have already implemented the relation $\left(s_{AA}^{-1}\right)^T+2\imath g_{AA}=0$ which remains the same as in (3.13); that is why s_{AA} is already elliminated from (3.22). This equation is nonlinear and potentially has many solutions. If we again take the fully replica-symmetric ansatz, we can easily solve for the trace as in (3.13)–(3.14). When everything is said and done the outcome is

$$
s=\pm\frac{2\imath}{t}I\otimes E,\ \ g=\pm tI\otimes E \tag{3.23}
$$

$$
S_{LLL'L'}=S_{LLR'R'}=S_{RRL'L'}=S_{RRR'R'}=\frac{4\imath}{L^{2N-4}}I \tag{3.24}
$$

$$
G_{LLL'L'}=G_{LLR'R'}=G_{RRL'L'}=G_{RRR'R'}=\frac{L^{2N-4}}{8}I. \tag{3.25}
$$

$$
W_4|_{\mathrm{on-shell}}\sim 4(N^2-N)\log L+2N\log 2-\sqrt{2N}/L^N. \tag{3.26}
$$

The remaining components of S and G (those not listed in (3.25)) are zero. The first and second term in $W_4|_{\mathrm{on-shell}}$ come from S and G fields; the third term comes from s and g. Although the first term clearly dominates over the others, and the last term is negligible, we have deliberately written all three to emphasize the different contributions. We notice the following:

1. The solution for W_4 factorizes at leading order, as we have $W_4\sim 4W_1$ on-shell (compare to (3.9)). What is more, it factorizes nontrivially, i.e. the expression for W_4 in (3.21) does not consist of four copies of W_1 (3.7). The leading contribution to (3.26) comes roughly from the $\log\det S$ term in (3.21) and involves fields of the form $a_L^\dagger a_L a_{L'}^\dagger a_{L'}$ and similar, whereas the dominant contribution to W_1 comes from the term $\log\det s$ with fields of the form $a_L^\dagger a_L$.

2. There is no simple "wormhole" contribution, mixing the left and right copy, i.e. no fields g_{LR}, only g_{LL} and g_{RR}. As we already said, this is so because there is no linear source coupling to a. However, since the fields $G_{AAB'B'}$ (or explicitly $G_{LLR'R'}$ and $G_{RRL'L'}$) are nonzero, this solution can be called half-wormhole (HWH) in analogy

JHEP09(2022)136

with the terminology of [11], already accepted in [15, 17, 18, 21].[11] The half-wormhole thus dominates the action.

3. The local contribution in replica space, the one coming from g_{LL} and g_{RR} fields, is negligible compared to the HWH part. It is also non-factorizing; but as in other cases in the literature, the HWH restores factorization.

4. The solution is not self-averaging. To see this, we follow the criterion (1.2) and insert $\sqrt{\langle\lambda_\mu^2\rangle} = L_\mu$ into Z as given in (3.3). Since now we deal with a free quadratic action (having fixed $\lambda_\mu \mapsto L_\mu$) the n-replica partition function is easily found for any n: $W_n = nN \log L$; specifically, for $n = 4$, we have $W_4 \sim 4N \log L$. This does not capture the leading term in (3.26).

We can confirm our findings numerically by computing the full action for a numerically generated ensemble of quenched matrices A_μ. In figure 1 we plot the numerical realization and the estimate (3.26) in blue and red respectively. We also plot the quadraple value of the single-replica action W_1 for comparison (black dashed). The analytical estimate is very good and the factorization near-perfect, as the black dashed curve almost falls on top of the blue and red one.

So while the action of the two-replica system W_2 does not factorize, the four-replica action W_4 has a HWH saddle point which restores factorization. In our view, the most interesting aspect of this is that the four-replica system *factorizes nontrivially*: the effective action (3.26) is dominated by nonlocal terms, half-wormholes; it does not merely consist of four copies of W_1. Since our system contains explicit quenched disorder, it is no big surprise that in general $\langle Z^2\rangle \neq \langle Z\rangle^2$ (this is the same situation as for the SYK model). But the restoration of the factorization in the sense that $\langle Z^4\rangle = \langle Z\rangle^4$ was not *a priori* expected, nor the fact that this happens already in the collective fields formalism at leading order (disregarding an infinite series of higher-order terms in $\langle Z^4\rangle$). Even though we look directly at the (discretized) string theory action, which includes gravity, there is no obvious geometric interpretation of the half-wormhole. This is not much of a surprise as the IKKT model should capture the nonperturbative stringy effects, far beyond general relativity and geometry.

3.4 Fermionic contribution to D-string partition functions

Now we turn on also the fermionic fluctuations ψ_α (while still putting the background fermions Ψ_α to zero). In addition to the collective fields introduced earlier, we now need also the fermion bilinear $\gamma_{AB} = \bar{\psi}_A\psi_B$, with $A, B \in \{L, R\}$; in order to implement the constraint that defines γ_{AB} we also have to introduce another auxiliary field called σ_{AB} (analogous to s_{AA}). Actually, the fermionic part may be more familiar to the reader as fermionic collective fields were investigated in the literature in much detail, mainly in the

[11] In fact, the name is not a very fortunate one, as there is nothing "halved" here; it a product of replicas just like a wormhole, only of different replicas. A more descriptive term would be "higher-order wormholes" but we will not attempt to change an already established name.

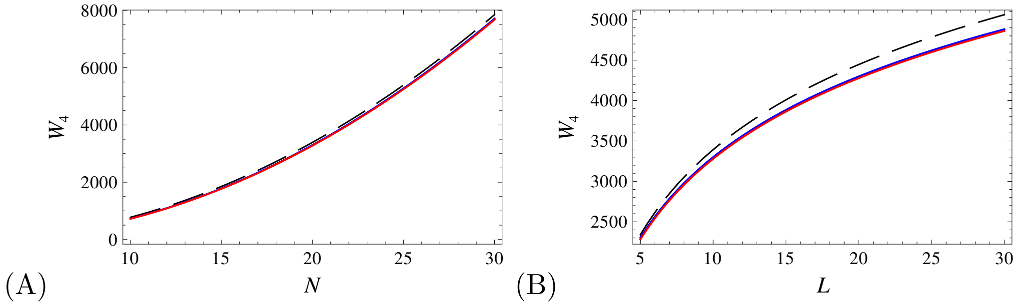

Figure 1. Effective action W_4 as a function of the matrix size N (A) and compactification radius L (B) from analytical (red) and numerical (blue) calculations; the curves almost coincide. For reference we also show the single-replica solution $4W_1$ (black dashed) which is nearly equal to W_4, confirming the factorizing property.

context of SYK model [12, 17, 20]. We can first write $\langle Z\rangle$:

$$\begin{aligned}\langle Z\rangle=&\int D[\psi]\int D[\bar\psi]\int D[a]\int D[g]\int D[\gamma]\int\frac{D[s]}{2^N\pi^{N^2}}\int\frac{D[\sigma]}{2^N\pi^{N^2}}\int d\lambda\\&\times\exp\mathrm{Tr}\left(-\frac{1}{2}aP^2a-\bar\psi_\alpha \not{P}\psi_\alpha\right)\\&\times\exp\left[-\imath\mathrm{Tr}\left(s^\dagger\left(g-a^\dagger a\right)+\sigma^\dagger\left(\gamma-\frac{1}{N}\bar\psi_\alpha\psi_\alpha\right)\right)-V(g)+V(a^\dagger a)-\Gamma(\gamma)+\Gamma\left(\frac{1}{N}\bar\psi_\alpha\psi_\alpha\right)\right]\\=&\int D[g]\int D[\gamma]\int\frac{D[s]}{2^N\pi^{N^2}}\int\frac{D[\sigma]}{2^N\pi^{N^2}}e^{-\imath\mathrm{Tr}\left(s^\dagger g+\sigma^\dagger\gamma\right)-V(g)-\Gamma(\gamma)+\mathrm{Tr}\left(\gamma s^{-1}\gamma\right)+\log\det\sigma},\end{aligned}\tag{3.27}$$

where we have introduced the self-interaction potential $\Gamma(\gamma)=(L^2/4)\gamma^2$ for the collective fermionic fields analogously to $V(g)$ for the boson. At quadratic order $O(g^2+\gamma^2)$ there is no interaction between the Bose and Fermi sectors. At the quartic level they couple, through the term $\gamma s^{-1}\gamma$ which comes from the term $\bar\psi\not{d}\psi$ in the original action. Defining the negative exponent in (3.27) as W_1, we will now compare W_1 to the four-replica effective action W_4. Let us first calculate W_1. The equations of motion read

$$\begin{aligned}&\frac{2}{L^{2N-2}}I+\imath s=0,\ \imath g+\frac{1}{2}\left(s^{-1}\right)^T+\gamma s^{-2}\gamma=0\ \Rightarrow\ s=\frac{2\imath}{L^{2N-2}}I\otimes E,\ g=\frac{L^{2N-2}}{4}I\otimes E\\&\imath\sigma+\frac{L^2}{2}\gamma-\frac{1}{4}\left(\gamma s^{-1}+s^{-1}\gamma\right)=0,\ \imath\gamma+\left(\sigma^{-1}\right)^T=0\ \Rightarrow\ \sigma=\imath\frac{L^{N-1}}{\sqrt2}I,\ \gamma=\frac{\sqrt2}{L^{N-1}}I.\end{aligned}\tag{3.28}$$

This yields the same solution for the bosonic fields g and s as before, since their coupling to the fermions (the last term in the first line of (3.28)) only contributes a correction which goes to zero for $L,N\to\infty$; in fact, the γ-s coupling does not influence the fermionic solution either at large L,N. The on-shell action is just the sum of the bosonic term (3.9) and the fermionic term from (3.28):

$$W_1|_{\text{on-shell}}=\left[(N^2-N)\log L+\frac{N}{2}\log 2\right]+\left[N\log L-\frac{N}{2}\log 2\right]=N^2\log L.\tag{3.29}$$

Now we derive the four-replica action W_4. In order to do this, we need to write first the two-replica function. Just like for bosons, it separates into the averaging and non-averaging

JHEP09(2022)136

part:

$$
\begin{aligned}
Z^2 &= \int \frac{D[s]}{2^N \pi^{N^2}} \int \frac{D[\sigma]}{2^N \pi^{N^2}} \chi(s,\sigma;g,\gamma)\Phi(s,\sigma) \\
\chi(s,\sigma;g,\gamma) &= \int D[g] \int D[\gamma] \exp\left[-\imath \mathrm{Tr}\left(s^\dagger_{AA} g_{AA} + \sigma^\dagger_{AB}\gamma_{AB}\right) - V\left(g_{AA}\right) - \Gamma(\gamma_{AB})\right] \\
\Phi(s,\sigma) &= \int D[a] \int D[\psi_A] \int D[\bar\psi_A] \exp\left[\imath \mathrm{Tr}\left(s^\dagger_{AA} a^\dagger_A a_A\right) - \frac{1}{2}\mathrm{Tr} a^\dagger_A P^2 a_A + V\left(a^\dagger_A a_A\right)\right] \\
&\quad \times \exp\left[\imath \mathrm{Tr}\left(\sigma^\dagger_{AA} \frac{\bar\psi_A \psi_A}{N}\right) - \mathrm{Tr}\bar\psi_A \not{P} \psi_A + \Gamma\left(\frac{\bar\psi_A\psi_A}{N}\right) - \mathrm{Tr}\bar\psi_A \not{a}_A \psi_A\right].
\end{aligned}
\tag{3.30}
$$

In order to test the factorization property we find $\langle \Phi^2 \rangle$ through the steps similar to those in eq. (3.19).

$$
\begin{aligned}
\langle \Phi^2(s) \rangle &= \int D[a_A] \int D[a_{A'}] \int d\bar\psi_A \int d\psi_A \int d\bar\psi_{B'} \int d\psi_{B'} \int \frac{dS}{2^N\pi^{N^2}} \int dG \\
&\times \exp\left[\imath \mathrm{Tr}\left(s^\dagger_{AA} a^\dagger_A a_A + s^\dagger_{A'A'} a^\dagger_{A'} a_{A'} + S^\dagger_{AAB'B'} a^\dagger_A a_A a^\dagger_{B'} a_{B'} + \sigma^\dagger_{AA}\frac{\bar\psi_A\psi_A}{N}\right)\right] \\
&\times \exp\left[V\left(a^\dagger_A a_A\right) + V\left(a^\dagger_{A'} a_{A'}\right) + \mathcal{V}\left(a^\dagger_A a_A, a^\dagger_{B'} a_{B'}\right) + \Gamma\left(\frac{\bar\psi_A\psi_A}{N}\right) - \mathrm{Tr}\bar\psi_A \not{a}_A \psi_A\right] \\
&\times \exp\left[-\imath \mathrm{Tr}\left(S^\dagger_{AAB'B'} G_{AAB'B'}\right) - \mathcal{V}\left(g_{AA}, g_{B'B'}; G_{AAB'B'}\right)\right] = \int \frac{dS}{2^N \pi^{N^2}} \int D[G] e^{-\tilde W_f}.
\end{aligned}
\tag{3.31}
$$

The final step is to integrate out the original fields. The bosonic integrals result in determinants and traces, as we have already seen, and the fermionic integrals are expressed in terms of Pfaffians, i.e. polynomials. Thanks to this the fermionic bilinear can contain any combination of L, L', R and R', giving rise to wormholes as we shall see. The effective action is finally:

$$
\begin{aligned}
W_f &= \imath \mathrm{Tr}\left(s^\dagger_{AA} g_{AA} + \sigma^\dagger_{AB}\gamma_{AB}\right) + V + \tilde W_f \\
\tilde W_f &= \frac{1}{2}\log\det\left(s_{AB} s_{A'B'} S_{AAB'B'}\right) - \imath \mathrm{Tr}\left(S^\dagger_{AAB'B'} G_{AAB'B'}\right) + \mathcal{V} + V_f \\
\mathcal{V}_f &= \frac{L^2}{4}\mathrm{Tr}\left(\gamma^2_{AB'}\right) - \mathrm{Tr}\left(\gamma_{AA} s^{-1}_{AA} \gamma_{AA}\right) - \log \mathrm{Tr}\left(\sigma_{LR}\sigma_{L'R'} - \sigma_{LL'}\sigma_{RR'} + \sigma_{LR'}\sigma_{RL'}\right),
\end{aligned}
\tag{3.32}
$$

and the bosonic potentials V and $\mathcal{V}$ are given in (3.12) and (3.21) respectively. The saddle-point equations (3.22) now get additional terms and two additional equations from fermionic contributions:

$$
\begin{aligned}
&\frac{1}{2}\left(s^{-1}_{AA}\right)^T + \imath g_{AA} + \gamma_{AA} s^{-2}_{AA}\gamma_{AA} = 0 \\
&\imath s_{AA} + \frac{2I}{L^{2N-2}} + \frac{8I}{L^{2N-4}}\mathrm{Tr}(g_{LL} + g_{RR}) - \frac{8}{L^{2N-4}}(g_{LL}+g_{RR})^T - \frac{8I}{L^{2N-4}}\mathrm{Tr} g_{B'B'} = 0 \\
&\frac{1}{2}\left(S^{-1}_{AAB'B'}\right)^T + \imath G_{AAB'B'} = 0, \quad -\imath S_{AAB'B'} - \frac{4}{L^{2N-4}} I = 0 \\
&\imath \sigma_{AB} + \frac{L^2}{2}\gamma_{AB} - \frac{1}{4}\{\gamma_{AB}, s^{-1}_{AB}\}\delta_{AB} = 0 \\
&\imath \gamma_{AB} - \sigma_{A'B'}\left(\sigma_{CD}\sigma_{C'D'} - \sigma_{CC'}\sigma_{DD'} + \sigma_{CD'}\sigma_{DC'}\right)^{-1} = 0.
\end{aligned}
\tag{3.33}
$$

JHEP09(2022)136

This Bose-Fermi saddle-point system has a rich structure. We notice the following properties:

1. The collective field γ_{AB} can now couple different replicas already at the quadratic level. This is simply a consequence of the fact that fermions couple linearly to the quenched bosonic degrees of freedom A_μ. We would have obtained the same situation for the bosonic fields if we coupled them linearly to a source. The Pfaffian obtained upon integrating out ψ_α fields has a combinatorial structure which allows the breaking of replica symmetry, i.e. some σ_{AB} may be zero and some nonzero.

2. One solution is obtained by setting $\sigma_{LR} = \sigma_{L'R'} \equiv \sigma \neq 0$ while all other components are zero. This yields[12]

$$\begin{aligned} \sigma_{LR} = \sigma_{L'R'} &= \frac{L^{N-1}}{\sqrt{2}}, \qquad \gamma_{LR} = \gamma_{L'R'} = -\imath\frac{\sqrt{2}}{L^{N-1}} \\ W_4|_{\text{on-shell}} &= 4N^2 \log L + 2N. \end{aligned} \tag{3.34}$$

This is a wormhole (WH), coupling the L and R copies. The term $2N$ spoils the factorization (compare to (3.29)).

3. We can have $\sigma_{LR'} = \sigma_{RL'} \equiv \sigma \neq 0$ while the other components of σ_{AB} are zero. This is a half-wormhole. This solution reads

$$\begin{aligned} \sigma_{LR'} = \sigma_{RL'} &= \frac{L^{N-1}}{\sqrt{2}}, \qquad \gamma_{LR'} = \gamma_{RL'} = -\imath\frac{\sqrt{2}}{L^{N-1}} \\ W_4|_{\text{on-shell}} &= 4N^2 \log L + N. \end{aligned} \tag{3.35}$$

The effective action is lower than (3.34) by N so this solution is thermodynamically preferrable compared to the wormhole, and also the mismatch from factorization is smaller (N compared to $2N$).

4. The third possibility is $\sigma_{LL'} = \sigma_{RR'} \equiv \sigma$ as the only nonzero component. This solution acquires a minus sign in the logarithm of the Pfaffian (the last term in (3.32)) hence it has a nonzero phase:

$$\begin{aligned} \sigma_{LL'} = \sigma_{RR'} &= \imath\frac{L^{N-1}}{\sqrt{2}}, \qquad \gamma_{LL'} = \gamma_{RR'} = \frac{\sqrt{2}}{L^{N-1}} \\ W_4|_{\text{on-shell}} &= 4N^2 \log L + \imath\pi. \end{aligned} \tag{3.36}$$

This is a half-wormhole but inequivalent to the half-wormhole from the previous point; it breaks the phase symmetry and has lower (real part of) free energy than the previous solutions. Apart from the phase factor, it is a factorizing solution, as we have $\Re W_4 \sim 4N^2 \log L = 4W_1$.

[12] In all examples we consider, the solutions for the bosonic fields g, s, G, S do not change at leading order in the presence of fermions, hence we do not write the expressions for them again.

5. The last possibility is $\sigma_{LL} = \sigma_{RR} \equiv \sigma$ as the only nonzero component. This solution contributes to the last equation in (3.33) with the opposite sign, and thus leads to a different saddle point:

$$\sigma_{LL} = \sigma_{RR} = \frac{L^{N/2-2}}{2^{7/4}N^{1/4}}e^{-\imath\pi/4}, \ \gamma_{LL} = \gamma_{RR} = \frac{2^{7/4}N^{1/4}}{L^{N/2-2}}e^{-\imath\pi/4} \tag{3.37}$$

$$W_4|_{\text{on-shell}} = 4(N^2 + N/2)\log L + N - \frac{1}{4}e^{-\imath\pi/4}. \tag{3.38}$$

This saddle point would not exist for Majorana fermions (as in the SYK model) since for Majoranas $\bar{\psi} = \psi$ and thus $\gamma_{LL} = \langle\psi\psi\rangle = 0$. It is a strongly suppressed solution as its on-shell action is by $\sim 2N\log L$ larger than the real part of W_4 of all previous solutions (3.34), (3.35), (3.36).

This situation is more akin to the SYK and similar field theory models in the literature than the purely bosonic case — several competing solutions of the local (3.37), wormhole (3.34), and half-wormhole type (3.35), (3.36). There may be more general solutions with less symmetry than the ones we found analytically above. Such solutions are hard to find, and we did not explore them in detail. In figure 2 we plot the landscape of the real part of the action for varying σ_{LR} and $\sigma_{L'R}$ (fixing the remaining couplings). The outcome is invariant to the sign change of σ which was to be expected from the effective action. The trivial vacuum $\sigma_{LR} = \sigma_{L'R} = \sigma_{LR'} = 0$ is one solution, but we also have a nontrivial solution with all WH and HWH couplings nonzero. It seems however that the latter is never the global minimum, i.e. it is a false vacuum. So we fail to find any thermodynamically stable solutions other than (3.34)–(3.37), but we still have no proof that they never exist.

Looking at the on-shell actions (3.34)–(3.38), we see that the dominant (least-action) solution, the LL' HWH configuration, restores the factorization. It is true though that fermionic contributions are always subleading to the $4N^2\log L$ term coming from the bosonic determinant, so one could also argue that the factorization is always roughly satisfied; but precisely the *dominant* fermionic saddle point, i.e. the true vacuum satisfies it also at the next order in L, N. This strongly suggests that the physical dynamics indeed tends to "know" about the factorization. The factorization remains nontrivial, as the same dominant term $\log\det S$ from the bosonic action still gives the main contribution. And — unlike the findings for the SYK model in [12] — *none of our solutions is self-averaging.* The total action for the fluctuations a_μ and ψ_α evaluated for $\lambda = L$ is easily found from (2.8) to be zero — the Bose and Fermi terms exactly cancel out each other, as it has to happen for a BPS configuration (the averaging over the quenched background A_μ spoils the supersymmetry and that is why the averaged on-shell actions we find are always nonzero; but expanding about fixed BPS solution for A_μ without averaging keeps the BPS property).

4 Replicas and factorization for a pair of interacting D-strings

We have found that for a single D-string the factorization of the partition function, broken for two replicas, is always re-established with four replicas essentially at the mean field level, and in a nontrivial way. A possible physical interpretation of this fact is simply

JHEP09(2022)136

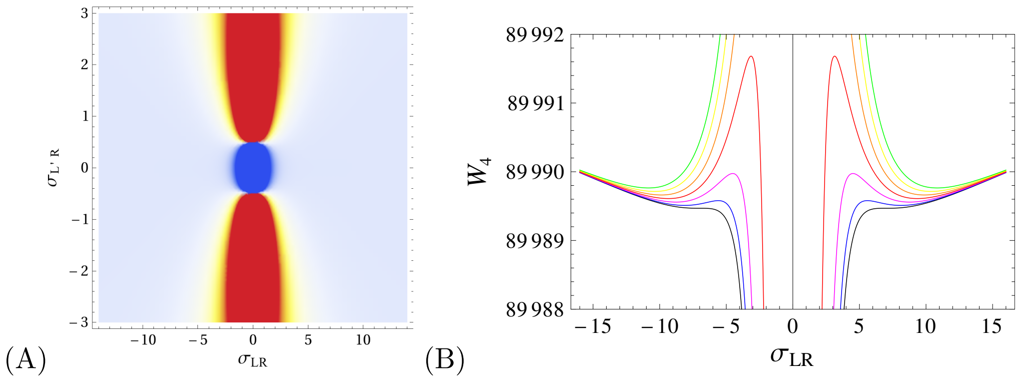

Figure 2. (A) Real part of the effective action W_4 as a function of the wormhole coupling σ_{LR} and the half-wormhole coupling $\sigma_{L'R}$, for $\sigma_{LL} = \sigma_{RR} = 0.3$ and $\sigma_{LL'} = \sigma_{RR'} = 0.5\imath$. The magnitude of W_4 is encoded by the color map, from blue (low) to red (high). The global minimum is the trivial solution $\sigma_{LR} = \sigma_{L'R} = 0$ (the blue area in the center). There is also a line of shallow local minima for nonzero WH and HWH couplings, which is easier to see along one-dimensional slices (B), where we plot $W_4(\sigma_{LR})$ for $\sigma_{L'R} = 0.1$, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7 (black, blue, magenta, red, orange, yellow, green). But the true vacuum, in the presence of nonzero LL and LL' couplings, remains the one with no WH (LR) or additional HWH ($L'R$) contributions (there is already one HWH coupling LL').

that n noninteracting D-strings form an ergodic system, i.e. averaging the dynamics over all the strings (computing their partition function Z_n) is equivalent to averaging a single string over the quenched degrees of freedom (computing the partition function $\langle Z_1 \rangle$). A consequence of this is that *if we have a BPS configuration, the conclusion from the previous section cannot change: as long as $F_{\mu\nu} = 0$ there is no difference between n copies of a single D-brane and a stack of n parallel D-branes.* Notice this remains true also in absence of background fermions (for $\Psi_\alpha = 0$) since it only hinges on $F_{\mu\nu} = 0$.

Now we consider a pair of strings with angle 2θ between them — this is the solution given in (2.5). For nonzero angle, this solution is non-BPS, the interaction between the strings is nonzero as it has a nonzero component of $F_{\mu\nu}$. Following eq. (2.10) and [30], the superoperators are found to be:

$$
\begin{aligned}
P_0 &= [q \otimes I_{2\times 2}, \cdot] = \begin{pmatrix} I_{2\times 2} \otimes \hat{Q} & 0 \\ 0 & I_{2\times 2} \otimes \hat{Q} \end{pmatrix} \\
P_1 &= \cos\theta \, [p \otimes I_{2\times 2}, \cdot] = \cos\theta \begin{pmatrix} I_{2\times 2} \otimes \hat{P} & 0 \\ 0 & I_{2\times 2} \otimes \hat{P} \end{pmatrix} \\
P_2 &= \frac{\ell}{2} \left[I_{N'\times N'} \otimes \sigma^3, \cdot \right] = \ell \begin{pmatrix} -\Pi_- \otimes I_{N'^2\times N'^2} & 0 \\ 0 & \Pi_+ \otimes I_{N'^2\times N'^2} \end{pmatrix} \\
P_3 &= \sin\theta \left[p \otimes \sigma^3, \cdot \right] = 2\sin\theta \begin{pmatrix} -\Pi_- \otimes \hat{Q} & 0 \\ 0 & \Pi_+ \otimes \hat{Q} \end{pmatrix} \\
F_{03} &= 2\omega \sin\theta \left[I_{N'\times N'} \otimes \sigma^3, \cdot \right] = 2\omega \sin\theta \begin{pmatrix} -\Pi_- \otimes I_{N'^2\times N'^2} & 0 \\ 0 & \Pi_+ \otimes I_{N'^2\times N'^2} \end{pmatrix},
\end{aligned}
\tag{4.1}
$$

and the other components are zero. For two strings, according to (2.5), the matrices A_μ

JHEP09(2022)136

and a_μ have the two-by-two block structure, hence the superoperators in (4.1) have the four-by-four block structure. We have denoted $N' \equiv N/2$ and $\Pi_\pm \equiv (I_{2\times 2} \pm \sigma^3)/2$. The superoperators $\hat{P} \equiv [p, \cdot]$ and $\hat{Q} \equiv [q, \cdot]$ are obtained by commuting with p and q and have the explicit form as given in (3.5). Now we follow the same path as in the previous section, so we will not repeat all the technical details. Unlike the previous section, we restore the spacetime indices and write e.g. $a^\dagger_\mu a_\mu$ and not $a^\dagger a$; this is because the different spacetime dimensions are not equivalent anymore as for a single string. The effective action reads

$$W_1 = \frac{1}{2}\log\det\left[\frac{I}{L^2} + \left(1+\cos^2\theta\right)\left(a^\dagger_\mu a_\mu - I\mathrm{Tr}a^\dagger_\mu a_\mu\right)\right] + \frac{\ell^2}{2}\sin^2\theta\cdot\mathrm{Tr}a^\dagger_\mu K^2 a_\mu$$
$$+ 2\omega\sin\theta\mathrm{Tr}\left(a^\dagger_0 K a_3 + a^\dagger_3 K a_0\right), \quad K = \begin{pmatrix} -\Pi_- & 0 \\ 0 & \Pi_+ \end{pmatrix}. \tag{4.2}$$

Introducing the collective fields, expanding the log det term, and integrating out the original variables we get

$$\begin{aligned} W_1 = \mathrm{Tr}\Bigg[&\frac{2\left(1+\cos^2\theta\right)}{L^{2N-2}}\left(g - a^\dagger_\mu a_\mu\right) + \frac{\ell^2}{2}\sin^2\theta\left(K^2 g - a^\dagger_\mu K^2 a_\mu\right) \\ &+ 2\omega\left(j - \sin\theta\left(a^\dagger_0 K a_3 + a^\dagger_3 K a_0\right)\right) \\ &+ \imath s^\dagger\left(g - a^\dagger_\mu a_\mu\right) + \imath\zeta^\dagger\left(j - \sin\theta\left(a^\dagger_0 K a_3 + a^\dagger_3 K a_0\right)\right)\Bigg] \\ = &\frac{1}{2}\log\det s^2\left(s^2 - \sin^2\theta K^2\zeta^2\right) \\ &+ \mathrm{Tr}\left[\frac{2\left(1+\cos^2\theta\right)}{L^{2N-2}}g + \frac{\ell^2}{2}\sin^2\theta K^2 g + 2\omega j + \imath s^\dagger g + \imath\zeta^\dagger j\right]. \end{aligned} \tag{4.3}$$

Here, j is the new current, equal (on-shell) to the bilinear originating from the string-string interaction, and ζ is the corresponding auxiliary field. The two-replica action (to quadratic order) is now

$$\begin{aligned} W_2 &= \frac{1}{2}\log\det s^2_{AA}\left(s^2_{AA} - \sin^2\theta K^2\zeta^2_{AA}\right) + V_2\left(g_{AA}\right) + \mathrm{Tr}\left[2\omega j_{AA} + \imath s^\dagger_{AA} g_{AA} + \imath\zeta^\dagger_{AA} j_{AA}\right] \\ V_2(g_{AA}) &= \mathrm{Tr}\left[\frac{2\left(1+\cos^2\theta\right)}{L^{2N-2}}g_{AA} + \frac{\ell^2}{2}\sin^2\theta K^2 g_{AA}\right]. \end{aligned} \tag{4.4}$$

Finally, the four-replica solution yields

$$\begin{aligned} W_4 &= \imath\mathrm{Tr}\left(s^\dagger_{AA}g_{AA} + \zeta^\dagger_{AA}j_{AA}\right) + V_2(g_{AA}) + 2\mathrm{Tr}\omega j_{AA} + \tilde{W}_4 \\ \tilde{W}_4 &= \frac{1}{2}\log\det\left[s^2_{AA}\left(s^2_{AA} - \sin^2\theta K^2\zeta^2_{AA}\right) + (A \mapsto A') + S^2_{AAB'B'}\left(S^2_{AAB'B'} - \sin^2\theta K^2\zeta^2_{AAB'B}\right)\right] \\ &\quad + \mathcal{V} \\ \mathcal{V} &= \frac{8}{L^{2N-4}}\mathrm{Tr}g_{AA}\mathrm{Tr}g_{B'B'} - \frac{4}{L^{2N-4}}\mathrm{Tr}G_{AAB'B'}. \end{aligned} \tag{4.5}$$

The solutions to the saddle-point equations for $W_{1,2,4}$ depend crucially on whether the strings are parallel ($\theta = 0$) or not ($\theta \neq 0$). We will therefore consider each case separately. Parallel strings with $\theta = 0$ and thus $j = 0$ reduce to four times the result of the single-matrix calculation from the previous section; therefore we have $W_1|_{on-shell} = 4N^2\log L$

(and analogously $W_4|_{on-shell} = 16N^2 \log L$).[13] This of course had to happen, as this case is BPS. The old conclusion thus remains, as we have anticipated at the beginning of this section: the partition functions factorize nontrivially.

For a non-BPS configuration the picture will actually turn out to be simpler. Let us first solve the saddle-point equations for W_1. The outcome is

$$s = \frac{2\imath}{L^{2N-2}}\left[(1+\cos^2\theta)I + \ell^2 \sin^2\theta K^2\right], \quad \zeta = 2\imath\omega I \tag{4.6}$$

$$j = \imath\frac{\sin^2\theta K^2\zeta}{K^2\zeta^2\sin^2\theta - s^3} \Rightarrow j \sim \frac{1}{2\omega}I \tag{4.7}$$

$$0 = \frac{s^2\left(2 + \imath gs - \sin^2\theta K^2 (I+gs)\zeta^2\right)}{s^3 - \sin^2\theta K^2 s\zeta^2} \Rightarrow g \sim \frac{I}{2L^{2N-2}(1+\cos\theta^2 + \ell^2\sin^2\theta K)} \tag{4.8}$$

This leads to the on-shell action (neglecting as usual the subleading terms):

$$\begin{aligned} W_1|_{\text{on-shell}} &\sim 2N^2 \log L + N^2 \log(4Lc_1c_2) + \ldots \\ c_1 &= 1 + \cos\theta^2 + \ell^2\sin^2\theta, \quad c_2 = \omega\sin\theta. \end{aligned} \tag{4.9}$$

It is easy to inspect W_1 and find that the leading contribution to the on-shell action value (4.9) comes from the term with $\log\det$ in (4.3). Now let us look at the two- and four-replica action. These lead to very cumbersome expressions which, however, do not present any principal difficulties so we do not write them out in full detail. The outcome reads

$$\begin{aligned} W_2|_{\text{on-shell}} &\sim 4N^2 \log L + 2N^2 \log(4Lc_1c_2) + \ldots \\ W_4|_{\text{on-shell}} &\sim 8N^2 \log L + 4N^2 \log(4Lc_1c_2) + \ldots \end{aligned} \tag{4.10}$$

One can also check that nothing changes qualitatively when $\ell = 0$, i.e. when the strings intersect. The limit of parallel strings is subtler: taking the limit $\theta \to 0$ directly does not make sense as it gives a divergence in W_1; instead one should go back to the BPS solution (3.8) for the collective fields, which brings the old result $W_1|_{on-shell} = 4N^2 \log L$.

Obviously, from (4.10), the action always factorizes. But the interesting part is that the terms which contribute at (leading) order $N^2 \log L$ all come from the same term $\log\det(s^2 \ldots)$, obtained from the bosonic determinant upon integrating out the microscopic variables a_μ. Therefore, *interacting systems have partition functions which factorize trivially.* This is in contrast to the non-interacting (and specifically BPS) configurations where $\langle Z^4\rangle$ factorizes nontrivially, from the sum of terms $\log\det s$ and $\log\det S$ (i.e., from both two-replica and four-replica couplings). An intuitive explanation would be the following: in BPS systems, the interactions (from gravity and from 2-form fields) precisely cancel out each other, but when multiple replicas are involved, there are combinations (like LR or $LLL'L'$) where this cancelation does not happen, so the structure of the effective action is different from that of a single replica (i.e. does not amount to four copies of a single replica). In the presence of interactions however, the dominant contribution to the free energy always comes from pairwise interactions: this is so for W_1, W_2, W_4 and for any W_n.

[13] The overall factor of four in these solutions compared to the solutions (3.9) and (3.26) comes from the fact that we now explicitly consider the four matrices A_μ, $\mu = 0\ldots 3$; in section 3 we were working with a single matrix for simplicity.

JHEP09(2022)136

4.1 Fermionic contributions

The fermionic contributions do not add anything fundamentally new to the picture, as they are subleading compared to the bosons. While the complete calculation is quite involved, it is easy to recognize the principle. Let us compute the fermionic contribution to the single-replica and two-replica actions. The fermionic contribution at leading order comes from the term $\bar{\psi}_\alpha \not{P} \psi_\alpha$ from (2.8). Using (4.1), integrating out the quenched degrees of freedom and then also the original fermionic fields ψ_α, we get the fermionic contribution in addition to the bosonic one from (4.2):

$$\begin{aligned} W_{1f} &= \imath \sigma^\dagger \gamma + \frac{L^2}{4} \gamma (K \otimes I_{N^2 \times N^2}) \gamma - \mathrm{Tr}(\gamma s^{-1} \gamma) - \log\det \left(\sigma - \frac{\ell}{2} (M \otimes I_{N^2 \times N^2}) \right) \\ M &= I_{2\times 2} \otimes \sigma_3 - \sigma_3 \otimes I_{2\times 2}, \quad K = 2 I_{4\times 4} + \ell \sin\theta M, \quad \gamma \sim \bar{\psi}_\alpha \psi_\alpha. \end{aligned} \tag{4.11}$$

We remind that the matrices are now of size $2N \times 2N$, and therefore the superoperators are of size $4N^2 \times 4N^2$, so they are naturally regarded as block matrices of size 4×4 with each block of size $N^2 \times N^2$. That is why we have the four-by-four matrices in the expressions above. The contribution proportional to ℓ comes from the gap, i.e. the finite separation of the D-strings. To actually compute the saddle point solutions to (4.11) is not easy, but if we look at the doubled system:

$$\begin{aligned} W_{2f} &= W_2 + \imath \mathrm{Tr} \sigma^\dagger_{AB} \gamma_{AB} + \frac{L^2}{4} \gamma_{AB'} (K \otimes I_{N^2 \times N^2}) \gamma_{AB'} - \mathrm{Tr} \left(\gamma_{AA} s^{-1}_{AA} \gamma_{AA} \right) \\ &\quad - \log \mathrm{Tr} \left(\tilde{\sigma}_{LR} \tilde{\sigma}_{L'R'} - \tilde{\sigma}_{LL'} \tilde{\sigma}_{RR'} + \tilde{\sigma}_{L'R} \tilde{\sigma}_{LR'} \right), \\ \tilde{\sigma}_{AB'} &\equiv \sigma_{AB'} - \frac{\ell}{2} M \otimes I_{N^2 \times N^2}, \end{aligned} \tag{4.12}$$

with W_2 being the bosonic contribution from (4.10), we see that the only difference with respect to the parallel strings is the transformation $\sigma \mapsto \sigma - (\ell/2) M \otimes I_{N^2 \times N^2}$. However, the determinant (and the trace) of the matrix σ does not change under this transformation. This can be checked directly as M is a degenerate matrix, with two zero eigenvalues. Therefore, the solutions and the on-shell values of the effective action stay the same as before. We still have the multiple choices (3.34)–(3.38), where (3.36), the true vacuum, actually leads to nontrivial factorization, but they all contribute only subleading terms to the bosonic action. Therefore, even though the outcome is mixed (trivial factorization for bosons plus nontrivial for fermions), the picture at leading order remains the same, with or without fermionic contributions — the multi-replica solution factorizes trivially at leading order.

To wrap up, the leading order collective field description (i.e. a mean field description) reproduces the factorization of partition functions of fluctuating D$_p$-branes (the number of dimensions p plays no role here), nontrivially for BPS-stabilized branes and trivially for interacting branes. We will try to make a bigger picture out of this fact in the final section.

5 Discussion and conclusions

The main outcome of our adventure can be summarized in the following way: (1) the restoration of factorization with $n \geq 4$ replicas (2) the nonfactorization of the two-replica

JHEP09(2022)136

system and the nontrivial factorization of the four-replica system for the noninteracting D-branes (3) the trivial factorization for any number of replicas in interacting D-brane systems (4) the absence of self-averaging. On one hand, the restoration of factorization is overall in line with the previous results from the literature for the SYK model [12, 15, 16, 23, 25] and the general expectation that our intuition should somehow survive so that the thermodynamic potentials of multiple independent copies of any system should just add up (i.e. the partition functions should just multiply). The points to ponder about are trivial vs. nontrivial factorization, the absence of self-averaging and above all the fact that we work with the IKKT model which is not a (fixed metric) field theory but is itself a quantum gravity system.

The absence of self-averaging, while in itself an important difference from the SYK and similar models, is maybe less surprising than it at first appears. Self-averaging means that the physics of the averaged model can be obtained as a small correction to a model with random fixed background (no matter which one!). That can only be true if the dynamics of the system is almost ergodic, i.e. if the system is strongly chaotic, perhaps only if it saturates the fast scrambling limit, like black holes and their dual field theories (such as the SYK model). Therefore, it is no surprise that self-averaging is never there for more general models. This has little to do with gravity in the IIB model; it only has to do with the fact that it is not (dual to) a black hole.[14] Thus we expect that a wide range of field theories will never be self-averaging (and the fact that our equations can be reinterpreted as the Eguchi-Kawai discretized Yang-Mills theory apparently corroborates that).

The trivial vs. nontrivial factorization is subtler. In the section 4 we have offered a somewhat handwaving explanation: the non-interacting nature of the BPS configurations is maintained by the precarious balance between different forces (in a single system), this balance is lost in multi-replica combinations and is only restored when the contributions from all combinations are included (this is the term $\log\det S_{AAB'B'}$). We have no proof that this will happen at all orders or for all BPS configurations. It is also not true that interacting systems necessarily have a trivial factorization. For example, the SYK model cannot in general be divided into mutually noninteracting subsystems, yet it factorizes nontrivially as found in [12, 23], in the sense that the factorization is restored roughly as $Z^2 \sim \mathrm{WH} + \mathrm{HWH}$, in other words *both* a wormhole and a half-wormhole contribute significantly. This is in contrast with our backgrounds, both BPS and non-BPS, where *only* half-wormholes contribute significantly and there is no WH+HWH solution. An additional caveat is the important insight of [23] that different collective field descriptions are possible, leading to different types of HWH solutions. It seems that the trivial factorization of interacting branes, with nonzero $F_{\mu\nu}$ terms, is really a consequence of the structure of interactions in the action of the IKKT model (eqs. (2.8)–(2.9)). It is an interesting task for future work to understand this better: is this a specific signature of theories with (quantum) gravity?

To answer that question, it would be useful to look at theories where we can directly construct interesting geometries, and interpret the disorder *geometrically*. This was done

[14] While the IKKT model should contain also the microscopic description of black holes, we don't know which background A_μ corresponds to that solution.

in [29], where a system of wrapped branes on complicated cycles within some compact manifold is represented as quiver quantum mechanics with a random superpotential. The basic strategy of our paper (modelling complicated solutions as effectively random, introducing the collective field formalism and then computing the effective action) is quite close to what was done in [29]. The authors find both replica-symmetric saddles and saddles which break the replica symmetry; the latter probably give rise to non-factorizing solutions whereas the former probably factorize.[15] In fact, as shown in [52], the same model in a different regime gives rise to ergodicity breaking already in classical dynamics, suggestive of glassy behavior. The appealing thing is that in this case the many local minima of the free energy have an obvious interpretation: these are geometries with multiple black holes, forming "molecules" [53] and a direct gravity analogue of structural glasses [54]. We hope to learn more about the factorization of leading saddles in such systems in our future work.

Speaking of (non)factorization, one technical point should be emphasized. We look mainly at averaged partition functions $\langle Z^n \rangle$ and find various saddle-point solutions to the effective actions obtained as $-\log\langle Z^n \rangle$. One could instead look directly at the average of the logarithm: $-\langle \log Z^n \rangle$. This latter variant is called the quenched free energy in the literature on disordered systems, whereas our effective actions W_n are called annealed free energies. The two differ precisely when the system is non-factorizing, and this criterion was used to explain the non-factorization due to replica wormholes in [47]. We use a slightly different criterion: direct comparison of annealed free energies for different numbers of replicas. For our purposes either of the two approaches (comparing $\log\langle Z^n \rangle$ to $\langle \log Z^n \rangle$ versus comparing $\log\langle Z^n \rangle$ among each other for different n) is suitable to check the factorization, although the difference between the two kinds of free energy is perhaps the more usual way in statistical physics.[16]

One might worry about the fact that we always disregard the terms which scale to zero as $L, N \to \infty$, i.e. the terms where L or N appears with a negative power. Indeed, we have not studied the behavior of such terms and there is no way to argue (except by performing additional calculations and checking explicitly) that these terms will obey the same factorization laws. However, even the starting ansatz for the background (with random Hermitian matrices) only makes sense for N large (otherwise we cannot speak of eigenvalue statistics), so the whole setup does not really address the regime of small N. Therefore, in the framework of our approach, the large-N results we find are sound.

In the context of very interesting findings of [27], one can also wonder if the D-branes of the IIB matrix model have a meaning analogous to the correlated bulk branes in 2D gravity models. In [27], such branes act as novel UV degrees of freedom which, when integrated out to arrive at an IR description, introduce nonlocalities in the gravity theory; in that paper it turns out that these nonlocalities save the factorization if tuned to a specific value. Essentially the same mechanism holds for our interacting D-branes — their interaction is also nonlocal and guarantees factorization for any number of replicas; when the (nonlocal) interaction is absent (in the BPS case), the factorization can be violated with $n = 2$ replicas,

[15] These are just our rough guesses; neither we nor [29] have explicitly checked the factorization of partition functions in this setup.

[16] We thank Souvik Banerjee for pointing out this issue.

and when restored it happens through a more complicated mechanism. We hope to reach a better understanding of this mechanism, and the relation to [27], in future work.

Finally, although we were motivated to look at (non)factorization by the puzzle arising from the replica wormholes in AdS/CFT (like most of the recent papers on the subject), this question is worth asking also on its own, as it is connected to dynamical and statistical properties of quantum gravity. Within our approach, we can in principle discuss also the matrix models of AdS geometries, with a CFT dual. One way would be to impose a background geometry along the lines of [55–57], by introducing a deformation of the action (2.1) to obtain matrices which satisfy the algebra of the generators of AdS isommetries. Another way is to consider a different matrix model, along the lines of [54, 58], which corresponds to D-particles in AdS. We discuss this briefly in appendix C, however a more complete treatment is again a task for future work.

Acknowledgments

I am grateful to K. Hashimoto, A. Tsuchiya, R. Meyer and S. Banerjee for inspiring discussions, to the anonymous referee for helpful comments and for pointing out additional relevant literature, and to the organizers of the conference "Strings and Fields 2021" for giving me the opportunity to present the first results of this work. This work has made use of the excellent Sci-Hub service. Work at the Institute of Physics is funded by the Ministry of Education, Science and Technological Development and by the Science Fund of the Republic of Serbia, under the Key2SM project (PROMIS program, Grant No. 6066160).

A Including the background fermions

Here we show that we do not lose anything by putting the background fermions to zero. Clearly this is a drastic change to the background as it destroys the supersymmetry and makes the on-shell action nonzero, but it should not strongly influence the factorization properties. The inclusion of background fermions merely shifts the action of the background to zero, but the background is frozen and not integrated over in the quenched regime. The dynamics and partition functions are determined by the fluctuations on top of the background, and these are not changed qualitatively when background fermions are included. Let us now show this explicitly, for a single degree of freedom (i.e., for noninteracting parallel D-strings). The outcome is that the fermionic fluctuations become even smaller (and for $\Psi = 0$, considered in the main text, they are already subleading to the bosonic ones), so our findings on factorization remain the same as those already found for bosonic fluctuations in sections 3 and 4.

Keeping nonzero Ψ and $\bar{\Psi}$, the action up to second order (we will not compute third- and fourth-order corrections) becomes:

$$S = S_1 + S_2 \tag{A.1}$$

$$S_1 = \mathrm{Tr}\left[\bar{\Psi}_\alpha \not{a} \Psi_\alpha + \bar{\Psi}_\alpha \not{P} \psi_\alpha + \bar{\psi}_\alpha \not{P} \Psi_\alpha\right] \tag{A.2}$$

$$S_2 = \mathrm{Tr}\left[-\frac{1}{4} a_\mu \left(P^2 \delta_{\mu\nu} + 2F_{\mu\nu}\right) a_\nu + \bar{\Psi}_\alpha \not{a} \psi_\alpha + \bar{\psi}_\alpha \not{a} \Psi_\alpha + \bar{\psi}_\alpha \not{P} \psi_\alpha\right]. \tag{A.3}$$

JHEP09(2022)136

The steps analogous to eq. (3.27) leading to the effective action still consist of elementary Gaussian integrations, but we have a few extra terms. Comparing to the last line in (3.27), we have two extra terms in the action at second order:

$$W_1 = \imath \mathrm{Tr}\left(s^\dagger g + \sigma^\dagger \gamma\right) + V(g) + \Gamma(\gamma) - \mathrm{Tr}\left(\gamma s^{-1} \gamma\right) + \log\det\sigma + \frac{1}{4}(g\gamma + \gamma g) + \frac{1}{2}\left(\gamma\sigma^{-1} + \sigma^{-1}\gamma\right), \tag{A.4}$$

the last two terms being absent when $\Psi = 0$. The equations of motion for s are unchanged compared to (3.28) but the others change:

$$\begin{aligned} &\frac{2}{L^{2N-2}} I + \imath s + \frac{1}{2}\gamma = 0, \qquad \imath g + \frac{1}{2}\left(s^{-1}\right)^T + \gamma s^{-2}\gamma = 0 \\ &\imath\sigma + \frac{L^2}{2}\gamma - \frac{1}{4}\left(\gamma s^{-1} + s^{-1}\gamma\right) + \frac{1}{2} g + \frac{1}{2}\sigma^{-1} = 0 \\ &\imath\gamma + \left(\sigma^{-1}\right)^T - \left(\gamma\sigma^{-2} + \sigma^{-2}\gamma\right) = 0. \end{aligned} \tag{A.5}$$

The relation between s and g is obviously unaffected, so we can express g in terms of s and insert in the remaining equations. The solution (ignoring the contributions which go to zero for large N and L) reads

$$s = \frac{4\imath}{L^{2N-2}} I \otimes E, \quad g = \frac{L^{2N-2}}{8} I \otimes E, \quad \sigma = -\frac{\imath L^{2N-4}}{8} I \otimes E, \quad \gamma = \frac{4}{L^{2N-2}} I \otimes E. \tag{A.6}$$

Inserting this into the action W_1, we find that both the first and the second extra term in the action are negligible since they scale as 1 and $1/L^{4N-4}$ respectively, the old fermionic terms are even more strongly subleading, and $\log\det\sigma$ doubles with respect to the on-shell value of W_1 found in eq. (3.8). The same doubling will occur also for a multiple-replica configuration, since the relation between σ and γ (the fourth equation in (3.33)) changes only trivially when we consider multiple copies of the system: all terms get the replica indices but there are no new terms — compare the mentioned equation in (3.33) and the third equation in (3.28).

One caveat remains: as mentioned in the main text, there is no guarantee that we have found all solutions. It is thus possible that solutions with less symmetry exist which change qualitatively in the presence of background fermions. Our numerical explorations (figure 2) suggest that this is unlikely, but clearly there is no rigorous proof.

B Hard vs. soft cutoff for matrix eigenvalues

In this appendix we will rederive some of our results assuming the uniform distribution of the eigenvalues of the background matrices A_μ within some compactification interval: $-L_\mu \leq \lambda_i^\mu \leq L_\mu$. This is the usual picture in the IKKT model [30, 34].[17] From now on we again equate all compactification radii, so that $L_\mu \equiv L$. We will see that, apart from

[17]Sometimes in the literature the factor $L_\mu/\sqrt{2\pi N}$ is pulled in front of the matrices A_μ, so the eigenvalues are distributed between $\pm\sqrt{2\pi L_\mu}$ and the commutators between A_μ's are just $\imath$ with no prefactors. We have adopted the conventions of [34] in this paper, where the eigenvalues are between $\pm L_\mu$ and the commutator acquires additional prefactors as in (2.5).

numerical factors of order unity, nothing changes compared to the Gaussian soft cutoff. The reason is essentially that we are looking at the quenched IKKT/large-N Yang-Mills model, expanding in small fluctuations a_μ, ψ_α and in large cutoff/compactification radius L; it is thanks to this quenched dynamics that the system is largely insensitive on the details of the eigenvalue distribution of the semiclassical background. If the matrices A_μ were annealed and not quenched, i.e. if we were to study the full dynamics of the background, this would not be true and we would have to be careful about the regulator. This point is explained in detail in [28] where the averaged theory depends crucially on the UV closure because in the setups of [28] both fast and slow degrees of freedom are dynamical.

Let us start from the defining expression for the single averaged partition function (3.3):

$$\langle Z\rangle = \int D[A_\mu]\int D[a_\mu]e^{-S(a_\mu;A_\mu)}\mathcal{P}(A_\mu) = \int D[a_\mu]\int_{-L_\mu}^{L_\mu} d^{2N}\lambda_\mu^i e^{-S\left(a_\mu;\lambda_\mu^i\right)}, \tag{B.1}$$

where we have now emphasized that the limits of integration for λ_μ^i are between $-L$ and L. Proceeding along the same lines as before, this yields the integral

$$\begin{aligned}\langle Z\rangle &= \int D[a_\mu]\int_{-L}^{L} d^{2N}\lambda_{\mu i}\Pi_{i<j}\left(\lambda_{\mu i}-\lambda_{\mu j}\right)^2\exp\left[-\frac{1}{4}a^\dagger_{\mu ij}\left(\lambda_{\mu i}^2+\lambda_{\mu j}^2\right)a_{\mu kl}\delta_{jk}\delta_{il}\right]\\ &= \int D[a_\mu]e^{-W_1}\end{aligned} \tag{B.2}$$

$$W_1 = \frac{1}{2}\sum_\mu \log\det\left(2a_\mu^\dagger a_\mu - 2I\mathrm{Tr}a_\mu^\dagger a_\mu\right) - \log\mathrm{Erf}\left(L\sqrt{\frac{1}{2}\mathrm{Tr}a_\mu^\dagger a_\mu}\right). \tag{B.3}$$

The error function Erf in the result is quite difficult to work with (we understand the error function of a matrix and in general functions of matrices in the usual way). But when we expand in a_μ small just like we did in the last line of (3.6) the result simplifies:

$$W_1 = a_\mu^\dagger a_\mu - \frac{4}{3}I\mathrm{Tr}a_\mu^\dagger a_\mu + O\left(\left(a_\mu^\dagger a_\mu\right)^4\right) \mapsto \frac{1}{2}\log\det s + \imath\mathrm{Tr}s^\dagger g + \frac{2}{3L^{2N-2}}\mathrm{Tr}g. \tag{B.4}$$

In the above equation, we have first performed a straightforward series expansion of the effective action W_1 in (B.3), and then we have introduced the collective field $g = a_\mu^\dagger a_\mu$ just like with the Gaussian cutoff. As we see, the additional error function terms, when expanded to quadratic order, merely change the coefficients in front of some of the terms. From here it is already obvious that the whole logic will remain the same as before; we believe there is no reason to repeat all the calculations again, as the factorization does not depend on the numerical coefficients (if say the term Trg has a different coefficient, the same coefficient will remain also in the two-replica term Trg_{AA}, and the same holds for the potentials $V(g_{AA})$).

Now we show the same result for the fermionic fluctuations. Starting from the basic expression (3.27) we can write

$$\begin{aligned}\langle Z\rangle &= \int D[a_\mu]\int D[\bar\psi_\alpha]\int D[\psi_\alpha]\int_{-L}^{L} d^{2N}\lambda_{\mu i}\Pi_{i<j}\left(\lambda_{\mu i}-\lambda_{\mu j}\right)^2\exp\left[-\frac{1}{2}a_\mu^\dagger P^2 a_\mu - \frac{1}{2}\bar\psi_\alpha \not{P}\psi_\alpha\right]\\ &= \int D[a_\mu]\int D[\bar\psi_\alpha]\int D[\psi_\alpha]e^{-W_1}\end{aligned}$$

$$e^{-W_1} = \sqrt{\frac{\pi}{2a_\mu^\dagger a_\mu}}e^{\frac{(\bar\psi_\alpha\psi_\alpha)^2}{2a_\mu^\dagger a_\mu}}\left[\mathrm{Erf}\left(\frac{La_\mu^\dagger a_\mu+\bar\psi_\alpha\psi_\alpha}{\sqrt{2a_\mu^\dagger a_\mu}}\right)-\mathrm{Erf}\left(\frac{-La_\mu^\dagger a_\mu+\bar\psi_\alpha\psi_\alpha}{\sqrt{2a_\mu^\dagger a_\mu}}\right)\right], \tag{B.5}$$

JHEP09(2022)136

where it is understood that expressions of the form $(\bar{\psi}_\alpha\psi_\alpha)^2/2a^\dagger_\mu a_\mu$ really mean $(2a^\dagger_\mu a_\mu)^{-1}(\bar{\psi}_\alpha\psi_\alpha)^2$, i.e. we divide by matrices the usual way, by multiplying by the matrix inverse. The result (B.5) is quite involved, but a series expansion again brings it to the form quite close to the Gaussian result (3.27):

$$W_1 = a^\dagger_\mu a_\mu - I\mathrm{Tr}a^\dagger_\mu a_\mu + \frac{L^2}{6}\left(\bar{\psi}_\alpha\psi_\alpha\right)^2 + O\left(\left(a^\dagger_\mu a_\mu\right)^4 + \left(\bar{\psi}_\alpha\psi_\alpha\right)^4\right) \Rightarrow$$
$$\Rightarrow W_1 = \imath\mathrm{Tr}\left(s^\dagger g + \sigma^\dagger\gamma\right) + V(g) + \Gamma(\gamma) + \log\det\sigma, \quad \Gamma(\gamma) = \frac{L^2}{6}\gamma^2, \tag{B.6}$$

and $V = V_2 + V_4$ remains unchanged from (3.12). The only change with respect to the Gaussian eigenvalue statistics is the coefficient in $\Gamma(\gamma)$, being $L^2/6$ instead of $L^2/4$. Therefore, the story remains the same: while the on-shell values of the actions W_n will change, the (non)factorization properties will not, as the coefficient change in Γ affects equally the actions W_n with any number n of replicas.

C Toward a holographic interpretation

So far, as we have mentioned in the Introduction, the connection of this work to AdS/CFT is only indirect, as the simple D-string configurations that we consider do not have an obvious CFT dual. We have also argued that it is nevertheless very relevant to know the factorization properties of theories with gravity, precisely in order to know if the restoration of factorization through half-wormholes despite the existence of non-factorizing wormhole saddles is a specific twist of holography or rather a generic phenomenon. But then again a realization of AdS replica wormholes, with a dual CFT, in the framework of IKKT model is clearly a worthy and exciting task. It is an open problem how to derive the holographic duality within the IKKT model, certainly deserving a separate work; here we just outline some ideas.

Recall first that the IIB matrix model at leading order coincides with the Eguchi-Kawai model of discretized 10-dimensional Yang-Mills theory [36]. Hence, there is a direct relation to a quantum field theory: all the partition functions we have calculated can be reinterpreted as field-theoretical partition functions of the Eguchi-Kawai system, and these should factorize for a set of non-interacting replicas.

The second thing to note is that the IKKT model, describing the type IIB string theory, should contain also the celebrated D3 brane background which gives rise to gravity on $\mathrm{AdS}_5 \times \mathbb{S}^5$, dual to the $\mathcal{N} = 4$ superconformal Yang-Mills. However, we do not know yet how to find such a solution explicitly in terms of matrices. A stack of parallel D3 branes is completely analogous to a pair of parallel strings that we consider, just with 4 matrices instead of 2. But the AdS geometry is not explicit in such a solution, as the starting action and its solutions (2.4) in terms of random Hermitian matrices still describe branes in flat space — the starting action already assumes a background, because the Schild action in the first place also describes strings on flat background. Therefore, we could in principle repeat the averaging over some set of fluctuating AdS configurations with our formalism but it is not known what this set should be.

JHEP09(2022)136

For this reason, there is a much studied idea [55, 56] to impose a geometric symmetry by deforming the model so that the matrices satisfy the commutation relations of the corresponding isommetries. For example, [57] realizes the (Lorentzian) AdS_2 geometry by imposing the structure of $SO(2,1)$ algebra. In our (Euclidean) case, the group is $SO(1,2)$ and the deformed model reads:[18]

$$S = -\mathrm{Tr}\left(\frac{1}{4}\left[X_\mu, X_\nu\right]^2 + c f_{\mu\nu\rho}\left[X_\mu, X_\nu\right] X_\rho\right), \tag{C.1}$$

where $f_{\mu\nu\rho}$ are the structure constants of $SO(1,2)$,[19] and c is the coupling constant. Classical equations of motion now read (putting again the background fermions to zero):

$$\left[X^\mu, \left[X_\mu, X_\nu\right]\right] + g f_{\nu\rho\sigma}\left[X_\rho, X_\sigma\right] = 0. \tag{C.2}$$

But the solution to this equation has to satisfy a nontrivial commutation relation, so it is more restricted than the IKKT brane solutions, where for $N \to \infty$ *any* commuting random matrices form a solution. Neverthless, for N large, the three commutation relations do not much reduce the space of solutions, and all but three eigenvalues can still be chosen at random from some interval (in other words, any reducible representation of $so(1,2)$, combined form arbitrary irreducible representations, is a solution). The fluctuating solutions of the form $X_\mu = A_\mu + a_\mu$ now yield the following structure of the quenched action up to second order, analogous to (2.8):

$$\begin{aligned} S_2 = \mathrm{Tr}\Big[&-\frac{1}{4} a_\mu \left(P^2\delta_{\mu\nu} + 2F_{\mu\nu}\right) a_\nu - \imath c f_{\mu\nu\rho}\Big(\left[a_\mu, A_\nu\right] A_\rho + \left[A_\mu, a_\nu\right] A_\rho + \left[A_\mu, A_\nu\right] a_\rho \\ &+ \left[a_\mu, a_\nu\right] A_\rho + \left[a_\mu, A_\nu\right] a_\rho + \left[A_\mu, a_\nu\right] a_\rho\Big)\Big]. \end{aligned} \tag{C.3}$$

We will not consider the third- and fourth-order terms. For the equations of motion (C.2), the superoperators are expressed as $P_\mu = A_\mu \otimes I + I \otimes A_\mu$ and $F_{\mu\nu} = -f_{\mu\nu\rho} A_\rho$. Making use of that and the antisymmetry of the structure constants, we arrive at the simplified action for the fluctuations:

$$S_2 = -\frac{1}{4} a_\mu \left[P^2\delta_{\mu\nu} + 2 f_{\mu\nu\rho}\left(A_\rho - 8cP_\rho - 4c\Pi A_\rho\right)\right] a_\nu - 2c f_{\mu\nu\rho} P_\nu A_\rho a_\mu, \tag{C.4}$$

where $\Pi \equiv \sigma_3 \otimes I - I \otimes \sigma_3$. Here we will be somewhat sloppy: integrating over the background A_μ requires us to know the statistical properties of eigenvalues and traces of the $so(1,2)$ representations. These are known from the literature (see e.g. the references in [56]) but the calculation goes beyond the scope of this short appendix. For large N the central limit theorems suggest that it still makes sense to assume that the eigenvalues are only

[18] It is not entirely clear if the fermionic part should receive any deformations, but it is hard to think of any other meaningful term involving $\bar{\Psi}, \Psi$ which satisfies the necessary symmetries. In any case, in this appendix we will ignore the fermions completely, as we only want to provide a roadmap to a hologrpahic setup, not to perform a precise calculation.

[19] Although the commutator of classical solutions was denoted by $f_{\mu\nu}$, there can be no confusion as the structure constants have three indices instead of two, and in this appendix we will not make use of the commutators $f_{\mu\nu}$ anyway.

JHEP09(2022)136

weakly correlated (which we approximate with totally uncorrelated), and limited by some cutoff function which we assume to be Gaussian, as we have shown in appendix B that the outcome is not sensitive to cutoff. In this case, paralleling the steps in the sections 3 and 4, we find the effective action for a single copy:

$$W_1 = \left(\frac{I}{2L^2} + 2a^\dagger_\mu a_\mu + 8c\left(a^\dagger_\mu + a_\mu\right) - 2I\mathrm{Tr}a^\dagger_\mu a_\mu\right) + \frac{1}{2L^2}\left(a^\dagger_\mu f_{\mu\nu\rho}\left(1 - 8c - 4c\Pi\right)a_\nu\right)^2 \mapsto$$
$$\mapsto W_1 + \imath s^\dagger s + \imath\zeta^\dagger_{\mu\nu}K_{\mu\nu}, \quad g \sim a^\dagger_\mu a_\mu, \ \ K_{\mu\nu} \sim a^\dagger_\mu a_\nu + a^\dagger_\nu a_\mu. \tag{C.5}$$

Therefore, because of the deformation, we have an additional set of collective fields $K_{\mu\nu}$ and the corresponding auxiliary fields $\zeta_{\mu\nu}$. Obviously, this is an interacting system, unlike the BPS parallel branes from section 3. Integrating out a_μ, we get the effective action in the form:

$$W_1 = \frac{3}{2}\log\det s^2 + \frac{3}{2}\log\det\zeta^2_{\mu\nu} + 3\tilde{V}_2(g) - \frac{1-8c}{2L^2}\mathrm{Tr}K_{\mu\nu} + \imath s^\dagger g + \imath\zeta^\dagger_{\mu\nu}K_{\mu\nu}$$
$$\tilde{V}_2(g) = \frac{2-32c^2}{L^{2N-2}}g. \tag{C.6}$$

In a system with four replicas, we get a result of the form

$$W_4 = 6\log\det s^2 + 6\log\det\zeta^2_{\mu\nu} + 3\tilde{V}_2(g_{AA}) + \tilde{W}_4\left(s_{AA}, g_{AA}, \zeta_{AAB'B;\mu\nu}, K_{AAB'B';\mu\nu}\right)$$
$$- \frac{1-8c}{2L^2}\mathrm{Tr}K_{AA;\mu\nu} + \imath s^\dagger_{AA}g_{AA} + \imath\zeta^\dagger_{AAB'B';\mu\nu}K_{AAB'B';\mu\nu}. \tag{C.7}$$

We do not need detailed expressions for $\tilde{W}_4$ (although tedious, they are straightforward to find). The key point is that, just like the findings (4.9)–(4.10) for a pair of interacting D-strings, the term $\log\det s^2$ always dominates, with the solutions $g = 3\imath s^{-1}$ and $s = -(2+32c^2)/L^{2N-2}\imath I$. Finally, this yields

$$W_n \sim -6nN^2\log L + 3nN\log(2+32c^2) + \ldots, \tag{C.8}$$

hence we again have trivial factorization.

A big caveat is in order: the interpretation of this and similar models is different from D-branes in the IKKT model; the bosonic matrices X_μ are not branes but quantized (non-commuting) coordinates [55, 56]. Besides, we have not properly derived the eigenvalue distribution for the background. The connection to the IIB matrix model is thus rather indirect. In fact, the main limitation in applying this scheme directly to holographic setups lies in the general lack of knowledge on how to encode AdS/CFT (i.e., the corresponding AdS backgrounds) in the IKKT action; the general idea and the formalism itself remains applicable for any background.

As a side note, another possible connection of D-brane matrix models to AdS backgrounds is the quiver matrix model of [58] (see also the references therein), where D-particles in AdS are described by a supersymmetric matrix action in $0+1$ dimension, i.e. with a time derivative, hence only the static limit fits into our formalism. This setup is less directly related to the IKKT model, so we do not have a precise idea on what the meaningful background solutions would look like and what the averaging would give.

JHEP09(2022)136

References

[1] A. Almheiri, T. Hartman, J. Maldacena, E. Shaghoulian and A. Tajdini, *Replica wormholes and the entropy of Hawking radiation*, *JHEP* **05** (2020) 013 [arXiv:1911.12333] [INSPIRE].

[2] G. Penington, S.H. Shenker, D. Stanford and Z. Yang, *Replica wormholes and the black hole interior*, *JHEP* **03** (2022) 205 [arXiv:1911.11977] [INSPIRE].

[3] A. Almheiri, T. Hartman, J. Maldacena, E. Shaghoulian and A. Tajdini, *The entropy of Hawking radiation*, *Rev. Mod. Phys.* **93** (2021) 035002 [arXiv:2006.06872] [INSPIRE].

[4] Y. Sekino and L. Susskind, *Fast scramblers*, *JHEP* **10** (2008) 065 [arXiv:0808.2096] [INSPIRE].

[5] S.H. Shenker and D. Stanford, *Stringy effects in scrambling*, *JHEP* **05** (2015) 132 [arXiv:1412.6087] [INSPIRE].

[6] J. Maldacena, S.H. Shenker and D. Stanford, *A bound on chaos*, *JHEP* **08** (2016) 106 [arXiv:1503.01409] [INSPIRE].

[7] F. Haake, *Quantum signatures of chaos*, Springer (2004).

[8] M.D. Mehta, *Random matrices*, Elsevier (2004).

[9] A. Altland and J. Sonner, *Late time physics of holographic quantum chaos*, *SciPost Phys.* **11** (2021) 034 [arXiv:2008.02271] [INSPIRE].

[10] A. Belin and J. de Boer, *Random statistics of OPE coefficients and Euclidean wormholes*, *Class. Quant. Grav.* **38** (2021) 164001 [arXiv:2006.05499] [INSPIRE].

[11] D. Stanford, *More quantum noise from wormholes*, arXiv:2008.08570 [INSPIRE].

[12] P. Saad, S.H. Shenker, D. Stanford and S. Yao, *Wormholes without averaging*, arXiv:2103.16754 [INSPIRE].

[13] P. Saad, S. Shenker and S. Yao, *Comments on wormholes and quantization*, arXiv:2107.13130 [INSPIRE].

[14] J. Maldacena and D. Stanford, *Comments on the Sachdev-Ye-Kitaev model*, *Phys. Rev. D* **94** (2016) 106002 [arXiv:1604.07818] [INSPIRE].

[15] B. Mukhametzhanov, *Half-wormhole in SYK with one time point*, *SciPost Phys.* **12** (2022) 029 [arXiv:2105.08207] [INSPIRE].

[16] B. Mukhametzhanov, *Factorization and complex couplings in SYK and in Matrix Models*, arXiv:2110.06221 [INSPIRE].

[17] S. Choudhury and K. Shirish, *Wormhole calculus without averaging from $O(N)^{q-1}$ tensor model*, *Phys. Rev. D* **105** (2022) 046002 [arXiv:2106.14886] [INSPIRE].

[18] A.M. García-García and V. Godet, *Half-wormholes in nearly AdS$_2$ holography*, *SciPost Phys.* **12** (2022) 135 [arXiv:2107.07720] [INSPIRE].

[19] A.M. García-García and V. Godet, *Euclidean wormhole in the Sachdev-Ye-Kitaev model*, *Phys. Rev. D* **103** (2021) 046014 [arXiv:2010.11633] [INSPIRE].

JHEP09(2022)136

[20] T.-G. Zhou, L. Pan, Y. Chen, P. Zhang and H. Zhai, *Disconnecting a traversable wormhole: Universal quench dynamics in random spin models*, *Phys. Rev. Res.* **3** (2021) 022024 [arXiv:2009.00277] [INSPIRE].

[21] K. Goto, Y. Kusuki, K. Tamaoka and T. Ugajin, *Product of random states and spatial (half-)wormholes*, *JHEP* **10** (2021) 205 [arXiv:2108.08308] [INSPIRE].

[22] B. Freivogel, D. Nikolakopoulou and A.F. Rotundo, *Wormholes from averaging over states*, arXiv:2105.12771 [INSPIRE].

[23] K. Goto, K. Suzuki and T. Ugajin, *Factorizing wormholes in a partially disorder-averaged SYK model*, *JHEP* **09** (2022) 069 [arXiv:2111.11705] [INSPIRE].

[24] F.S. Nogueira, S. Banerjee, M. Dorband, R. Meyer, J.v.d. Brink and J. Erdmenger, *Geometric phases distinguish entangled states in wormhole quantum mechanics*, *Phys. Rev. D* **105** (2022) L081903 [arXiv:2109.06190] [INSPIRE].

[25] S. Banerjee, M. Dorband, J. Erdmenger, R. Meyer and A.-L. Weigel, *Berry phases, wormholes and factorization in AdS/CFT*, *JHEP* **08** (2022) 162 [arXiv:2202.11717] [INSPIRE].

[26] M. Berkooz, N. Brukner, S.F. Ross and M. Watanabe, *Going beyond ER=EPR in the SYK model*, *JHEP* **08** (2022) 051 [arXiv:2202.11381] [INSPIRE].

[27] A. Blommaert, L.V. Iliesiu and J. Kruthoff, *Gravity factorized*, arXiv:2111.07863 [INSPIRE].

[28] J.J. Heckman, A.P. Turner and X. Yu, *Disorder averaging and its UV discontents*, *Phys. Rev. D* **105** (2022) 086021 [arXiv:2111.06404] [INSPIRE].

[29] D. Anninos, T. Anous and F. Denef, *Disordered quivers and cold horizons*, *JHEP* **12** (2016) 071 [arXiv:1603.00453] [INSPIRE].

[30] N. Ishibashi, H. Kawai, Y. Kitazawa and A. Tsuchiya, *A large-N reduced model as superstring*, *Nucl. Phys. B* **498** (1997) 467 [hep-th/9612115] [INSPIRE].

[31] H. Aoki, S. Iso, H. Kawai, Y. Kitazawa and T. Tada, *Space-time structures from type IIB matrix model*, *Prog. Theor. Phys.* **99** (1998) 713 [hep-th/9802085] [INSPIRE].

[32] I. Chepelev, Y. Makeenko and K. Zarembo, *Properties of D-branes in matrix model of IIB superstring*, *Phys. Lett. B* **400** (1998) 40.

[33] H. Aoki, S. Iso, H. Kawai, Y. Kitazawa, A. Tsuchiya and T. Tada, *IIB matrix model*, *Prog. Theor. Phys. Suppl.* **134** (1999) 47 [hep-th/9908038] [INSPIRE].

[34] K.L. Zarembo and Yu. M. Makeenko, *Vvedenie v matrichnye modeli superstrun* (in Russian), *Usp. Fiz. Nauk* **168** (1998) 3.

[35] B. Ydri, *Review of M(atrix)-theory, type IIB matrix model and matrix string theory*, arXiv:1708.00734 [INSPIRE].

[36] T. Eguchi and H. Kawai, *Reduction of dynamical degrees of freedom in the large-N gauge theory*, *Phys. Rev. Lett.* **48** (1982) 1063 [INSPIRE].

[37] H. Kawai and M. Sato, *Perturbative Vacua from IIB Matrix Model*, *Phys. Lett. B* **659** (2008) 712 [arXiv:0708.1732] [INSPIRE].

[38] H. Kawai, S. Kawamoto, T. Kuroki, T. Matsuo and S. Shinohara, *Mean field approximation of IIB matrix model and emergence of four-dimensional space-time*, *Nucl. Phys. B* **647** (2002) 153 [hep-th/0204240] [INSPIRE].

[39] S.-W. Kim, J. Nishimura and A. Tsuchiya, *Expanding (3+1)-dimensional universe from a Lorentzian matrix model for superstring theory in (9+1)-dimensions*, *Phys. Rev. Lett.* **108** (2012) 011601 [arXiv:1108.1540] [INSPIRE].

[40] J. Nishimura and A. Tsuchiya, *Complex Langevin analysis of the space-time structure in the Lorentzian type IIB matrix model*, *JHEP* **06** (2019) 077 [arXiv:1904.05919] [INSPIRE].

[41] F.R. Klinkhamer, *On the emergence of an expanding universe from a Lorentzian matrix model*, *PTEP* **2020** (2019) 103 [arXiv:1912.12229] [INSPIRE].

[42] F.R. Klinkhamer, *A first look at the bosonic master-field equation of the IIB matrix model*, *Int. J. Mod. Phys. D* **30** (2021) 2150105 [arXiv:2105.05831] [INSPIRE].

[43] N. Kawahara and J. Nishimura, *The large N reduction in matrix quantum mechanics: A bridge between BFSS and IKKT*, *JHEP* **09** (2005) 040 [hep-th/0505178] [INSPIRE].

[44] J.M. Maldacena and L. Maoz, *Wormholes in AdS*, *JHEP* **02** (2004) 053 [hep-th/0401024] [INSPIRE].

[45] Y. Chen, V. Gorbenko and J. Maldacena, *Bra-ket wormholes in gravitationally prepared states*, *JHEP* **02** (2021) 009 [arXiv:2007.16091] [INSPIRE].

[46] D. Marolf and H. Maxfield, *Transcending the ensemble: baby universes, spacetime wormholes, and the order and disorder of black hole information*, *JHEP* **08** (2020) 044 [arXiv:2002.08950] [INSPIRE].

[47] N. Engelhardt, S. Fischetti and A. Maloney, *Free energy from replica wormholes*, *Phys. Rev. D* **103** (2021) 046021 [arXiv:2007.07444] [INSPIRE].

[48] J.-M. Schlenker and E. Witten, *No ensemble averaging below the black hole threshold*, *JHEP* **07** (2022) 143 [arXiv:2202.01372] [INSPIRE].

[49] H. Verlinde, *Wormholes in quantum mechanics*, arXiv:2105.02129 [INSPIRE].

[50] J. Nishimura and G. Vernizzi, *Spontaneous breakdown of Lorentz invariance in IIB matrix model*, *JHEP* **04** (2000) 015 [hep-th/0003223] [INSPIRE].

[51] J. Nishimura and G. Vernizzi, *Brane world generated dynamically from string type IIB matrices*, *Phys. Rev. Lett.* **85** (2000) 4664 [hep-th/0007022] [INSPIRE].

[52] D. Anninos, T. Anous, F. Denef, G. Konstantinidis and E. Shaghoulian, *Supergoop dynamics*, *JHEP* **03** (2013) 081 [arXiv:1205.1060] [INSPIRE].

[53] D. Anninos, T. Anous, P. de Lange and G. Konstantinidis, *Conformal quivers and melting molecules*, *JHEP* **03** (2015) 066 [arXiv:1310.7929] [INSPIRE].

[54] D. Anninos, T. Anous, F. Denef and L. Peeters, *Holographic vitrification*, *JHEP* **04** (2015) 027 [arXiv:1309.0146] [INSPIRE].

[55] H. Steinacker, *Emergent geometry and gravity from matrix models: an introduction*, *Class. Quant. Grav.* **27** (2010) 133001 [arXiv:1003.4134] [INSPIRE].

[56] H. Steinacker, *Non-commutative geometry and matrix models*, *PoS* **QGQGS2011** (2011) 004 [arXiv:1109.5521] [INSPIRE].

[57] D. Jurman and H. Steinacker, *2D fuzzy anti-de Sitter space from matrix models*, *JHEP* **01** (2014) 100 [arXiv:1309.1598] [INSPIRE].

[58] T. Anous and C. Cogburn, *Mini-BFSS matrix model in silico*, *Phys. Rev. D* **100** (2019) 066023 [arXiv:1701.07511] [INSPIRE].