

JT gravity and the ensembles of random matrix theory

DOUGLAS STANFORD AND EDWARD WITTEN

We generalize the recently discovered relationship between JT gravity and double-scaled random matrix theory to the case that the boundary theory may have time-reversal symmetry and may have fermions with or without supersymmetry. The matching between variants of JT gravity and matrix ensembles depends on the assumed symmetries. Time-reversal symmetry in the boundary theory means that unorientable spacetimes must be considered in the bulk. In such a case, the partition function of JT gravity is still related to the volume of the moduli space of conformal structures, but this volume has a quantum correction and has to be computed using Reidemeister-Ray-Singer “torsion.” Presence of fermions in the boundary theory (and thus a symmetry $(-1)^F$) means that the bulk has a spin or pin structure. Supersymmetry in the boundary means that the bulk theory is associated to JT supergravity and is related to the volume of the moduli space of super Riemann surfaces rather than of ordinary Riemann surfaces. In all cases we match JT gravity or supergravity with an appropriate random matrix ensemble. All ten standard random matrix ensembles make an appearance – the three Dyson ensembles and the seven Altland-Zirnbauer ensembles. To facilitate the analysis, we extend to the other ensembles techniques that are most familiar in the case of the original Wigner-Dyson ensemble of hermitian matrices. We also generalize Mirzakhani’s recursion for the volumes of ordinary moduli space to the case of super Riemann surfaces.

1	Introduction	1476
2	Random matrix ensembles and bulk topological theories	1482
3	Torsion	1528
4	Loop equations	1577

5	Comparison to JT gravity and supergravity	1596
Appendix A	Volumes of symplectic supermanifolds	1620
Appendix B	Behavior under $N \rightarrow N + 4$	1635
Appendix C	Super-Schwarzian path integrals for the disk and trumpet	1637
Appendix D	Mirzakhani’s recursion relation and its superanalog	1642
Appendix E	Comments on nonperturbative effects	1663
Appendix F	Comments on the minimal string	1669
References		1671

1. Introduction

Jackiw-Teitelboim (JT) gravity [1, 2] is a simple theory of two-dimensional quantum gravity that describes rigid hyperbolic spaces. It was used as a model for $\text{AdS}_2/\text{CFT}_1$ in [3], and more generally it describes the low-energy dynamics of any near-extremal black hole.

The theory is so simple that it is almost trivial. On a closed Euclidean manifold Y , the path integral reduces to the volume of a finite-dimensional space – the moduli space of Riemann surfaces with the topology of Y . However, in most applications, it is interesting to consider JT gravity on a space Y with at least one asymptotic boundary. In this setting, one also has to do a path-integral over “wiggles” associated to each boundary [4–6].

These wiggles are governed by a solvable theory known as the “Schwarzian theory.” This provides a link between JT gravity and the Sachdev-Ye-Kitaev (SYK) model [7–9], which reduces at low energies to the same Schwarzian theory [8–10]. For a review, see [11]. This correspondence has motivated much recent work on JT gravity.¹

¹This has included many different derivations of the exact path integral [12–20], studies of the Hilbert space structure [18, 21–23], correlation functions [24–26], the physical symmetry charges [27], and the on-shell action [28]. Further work has included generalizations to include higher spin fields [29], supersymmetry [30], different topologies [23, 26, 31, 32], and de Sitter space [33, 34].

In [35] it was shown that the path integral of JT gravity on arbitrary orientable surfaces is computed by a Hermitian matrix integral, with the dictionary

$$(1.1) \quad Z_{\text{JT}}(\beta_1, \dots, \beta_n) \leftrightarrow \left\langle \text{Tr} e^{-\beta_1 H} \dots \text{Tr} e^{-\beta_n H} \right\rangle.$$

On the LHS, $Z_{\text{JT}}(\beta_1, \dots, \beta_n)$ is the JT gravity partition function with n asymptotic boundaries with regularized lengths β_1, \dots, β_n . On the RHS, the angle braces imply an average of H over an appropriate random matrix ensemble. The contribution of a particular topology to the path integral on the LHS coincides with the contribution at a given order in the “genus expansion” of the matrix integral on the RHS. The crucial fact underlying (1.1) is the correspondence [36] between Mirzakhani’s recursion relation for the volumes of moduli space [37] and Eynard-Orantin “topological recursion” [38] which gives the genus expansion of a Hermitian matrix integral [39].

In this paper, we will generalize this in multiple directions by studying JT gravity on unorientable surfaces, by including a sum over spin structures, and by studying $\mathcal{N} = 1$ supersymmetric JT gravity [30, 40–42]. We also allow for the possibility of including a topological field theory that weights crosscaps and/or spin structures in a particular fashion. All together, we will consider three bosonic theories, twelve theories with fermions and no supersymmetry, and ten supersymmetric theories. We will find evidence that all of these are related to matrix integrals.

The relevant matrix integrals are characterized by two pieces of data: first, one has to specify the “spectral curve,” or equivalently the leading approximation to the density of eigenvalues (generalizing the Wigner semi-circle). Second, one has to specify a discrete choice of one of ten different symmetry classes. These classes consist of the three Dyson β -ensembles [43] and the seven (α, β) -ensembles of Altland and Zirnbauer [44] (three of these were previously discussed by Verbaarschot and Zahed in [45, 46]).² Which symmetry class is appropriate depends on the algebra of the random matrix, the time-reversal operator \mathbb{T} (if present) and the symmetry $(-1)^F$ (if present). The genus expansion for each of these classes of matrix integrals can be analyzed efficiently using the loop equations [47] in their modern form, following [39].

²We will use the notation β for the Dyson index, reserving β for the inverse temperature.

The basic task of this paper is as follows. For each of the various bosonic, fermionic and supersymmetric bulk theories, we determine the two pieces of data described above, thus specifying a candidate matrix integral dual. Then we check that the bulk theory and the matrix integral agree at higher orders (in some cases these checks are limited, and in some cases they are to all orders). In more detail, to match the leading approximation to the density of eigenvalues, one has to compute the bulk partition function on the disk topology. To match the discrete choice of symmetry class, one has to match anomalies in the discrete symmetries \mathbb{T} and $(-1)^F$. Having fixed this data, to check agreement at higher orders, one can compare the JT path integral on some higher topology space with the prediction of the matrix integral derived using the loop equations.

We will find evidence that the three bosonic JT theories are dual to matrix integrals where the Hamiltonian H is drawn from one of the three Dyson β -ensembles. Similarly, the ten supersymmetric JT theories are dual to matrix integrals where $H = Q^2$ and the supercharge Q is drawn from one of the three Dyson ensembles, or from one of the seven Altland-Zirnbauer ensembles. For the twelve theories involving spin structures but no supersymmetry, we show how to reduce the problem to the three bosonic cases. A summary of the different cases is as follows.

Bosonic, orientable only: This was the case considered in [35]. The volumes of moduli space, and thus the JT path integral, are related to a $\beta = 2$ (GUE-like) matrix integral.

Bosonic, orientable + unorientable (two subcases): In bosonic JT gravity, it is natural to consider a sum over not necessarily orientable manifolds. In the context of holographic duality, this is appropriate if the boundary theory has time-reversal symmetry. There are two versions of the sum, depending on whether or not we include a factor of $(-1)^{n_c}$ where n_c is the number of crosscaps. These two choices amount to two different bulk theories, and they correspond to $\beta = 1$ (GOE-like) and $\beta = 4$ (GSE-like) matrix integrals. As it turns out, the volume of moduli space of a unorientable surface has a logarithmic divergence due to the contribution of small crosscaps. The corresponding matrix integrals also have a divergence in the relevant double-scaled limit, and they predict the correct formula for the moduli space measure associated to a crosscap. However, because of the divergence, we are not able to study arbitrary-genus correlators in these cases.

Fermionic but nonsupersymmetric (twelve subcases): In holographic duality, a bulk geometry should have a spin structure or a less familiar analog of this known as a pin^+ or pin^- structure if the boundary theory has fermions and, respectively, no time-reversal symmetry; time-reversal symmetry with

$T^2 = (-1)^F$ (in other words, T^2 anticommutes with elementary fermions); or time-reversal symmetry with $T^2 = 1$. Hence it is natural to consider JT gravity on an orientable manifold with a spin structure or on a not necessarily orientable manifold with a pin^+ or pin^- structure. The sum over spin, pin^+ , or pin^- structures can be enriched with a topological field theory, for which the number of possible choices is 2, 2, or 8. We match these $2 + 2 + 8 = 12$ theories with random matrix constructions. New random matrix ensembles are not needed; it suffices to consider the three Dyson ensembles together with the assumed global symmetries.

Supersymmetric, orientable only (two subcases): New matrix ensembles are needed if we incorporate not just spin structures but $\mathcal{N} = 1$ supersymmetry. In this case, JT gravity is replaced with JT supergravity, whose partition function is an integral over the moduli space of super Riemann surfaces. Restricting to orientable surfaces, there are two subcases because of the freedom to include a factor of $(-1)^\zeta$, where ζ is the mod 2 index, weighting odd spin structures with a minus sign. These two choices of bulk theory correspond to two particular symmetry classes of matrix ensemble. In the case where we weight all spin structures equally, the supercharge Q is drawn from a GUE-like ensemble with a leading distribution of eigenvalues that is nonzero everywhere on the real line. The loop equations for such a matrix integral imply that essentially all correlations vanish. Correspondence with JT supergravity then predicts that the volume of the moduli space of genus g super Riemann surfaces (summed over even and odd spin structures) vanishes. For the second bulk theory in which we weight by $(-1)^\zeta$, the right matrix ensemble is one in which Q is drawn from an Altland-Zirnbauer ensemble. The loop equations for this matrix integral predict a nontrivial recursion for the volumes of super moduli space weighted by $(-1)^\zeta$. We prove this directly using a generalization of Mirzakhani's approach in the bosonic case.

Supersymmetric, orientable + unorientable (eight subcases): In $\mathcal{N} = 1$ JT supergravity with time-reversal symmetry (which necessarily is of pin^- type), we have to include a sum over pin^- structures. This sum can be weighted by one of eight possible topological field theories, corresponding to the factor $e^{-i\pi N' \eta/2}$ where η is the η -invariant of Atiyah-Patodi-Singer [48], and N' is an integer mod 8. These cases can be matched to matrix integrals by comparing the anomalies in time-reversal T and $(-1)^F$. In six of these eight cases, there is a divergence from the integral over small crosscaps, as in the bosonic theory. Again, the crosscap measure is precisely matched by the corresponding matrix integral, but the divergence prevents us from going to higher orders. In the other two cases, the contribution of the crosscap is

finite. Surprisingly, for the two cases where the crosscap is finite, the matrix integral predicts that all higher-order contributions vanish, indicating a cancellation between volumes of moduli spaces of orientable and unorientable super-surfaces.

We will now discuss the plan of the paper. In **section two** we review the ten standard ensembles of random matrix theory, and match anomalies in $(-1)^F$ and \mathbb{T} symmetry in order to line them up with JT gravity and supergravity theories. A useful tool in this section is the SYK model. Variants of this theory have an approximate random matrix classification, as well as an approximate relation to JT gravity and JT supergravity. Since both of these relationships are approximate, SYK by itself does not imply an exact correspondence between JT gravity and random matrix theory. However, assuming that such a correspondence does exist, the SYK model can be used to match discrete choices on the two sides.

In this section, in addition to matching the anomalies, we show that for the twelve fermionic but not supersymmetric theories, agreement with random matrix theory follows from agreement in the three bosonic cases. To demonstrate this, we evaluate the sum over spin, pin^+ and pin^- structures weighted by appropriate topological field theories.

In **section three** we show how to compute the measure on the moduli space of Riemann surfaces or super Riemann surfaces, starting with the path integral of JT gravity or JT supergravity. These theories can be formulated as BF theories with gauge group $\text{SL}(2, \mathbb{R})$ or $\text{OSp}(1|2)$ [40, 49, 50]. In general, the partition function of a BF theory reduces to an integral over a moduli space of flat connections. On an orientable two-manifold, the appropriate measure on the moduli space is defined by classical formulas (for example, the Weil-Petersson measure on moduli space in the case $\text{SL}(2, \mathbb{R})$, as discussed in [35]), but on an unorientable surface, there is an important one-loop correction, which is given by a ratio of determinants known as the “torsion.” The role of torsion in what is now known as BF theory was originally shown in [51]. We discuss general features of the torsion and then show how to compute it for $\text{SL}(2, \mathbb{R})$ and $\text{OSp}(1|2)$. As an important special case, we compute explicitly the measure factor associated to the size modulus of a crosscap. It turns out that the resulting measure has been defined previously from a different point of view [52].

In **section four** we analyze the loop equations [47] of random matrix theory. The loop equations are a general tool for computing the genus expansion of correlation functions in random matrix ensembles. The method was

significantly streamlined in [39] for $\beta = 2$ ensembles. We explain this streamlining and imitate it for the general β -ensembles and the Altland-Zirnbauer (α, β) ensembles. Previous work on the β -ensembles includes [53–55].

Finally, in **section five** we compare the predictions of the loop equations to JT gravity and JT supergravity. In all cases, we are able in lowest order to match the crosscap measure computed from the torsion plus topological field theory with a calculation in random matrix theory. In higher orders, most of the cases present difficulties, either due to a divergent volume of moduli space, or due to the difficulty in evaluating the volume of supermoduli space by independent means. However, for the one supersymmetric case where the matrix integral predicts that the volumes are both finite and nonzero (orientable only, $(-1)^\zeta$ weighting), we use the loop equations to predict a simple recursion relation for the volumes of super moduli space, which we prove directly in appendix D by adapting Mirzakhani’s approach to super Riemann surfaces. This recursion relation has also been obtained in unpublished work by Norbury, who has studied the spectral curve relevant to JT supergravity from a different point of view [56], following earlier work on the related “Bessel” spectral curve [57, 58]. The derivation in appendix D is based on a super McShane identity, which in the prototypical case of a genus 1 surface with one puncture has been discovered independently by Y. Huang, R. Penner, and A. Zeitlin [59].

In appendix A, we describe a general formula for volumes of symplectic supermanifolds that makes contact between our statements and some of Norbury’s results. In appendix B, we explain a relationship between certain pairs of matrix ensembles, generalizing a known relation between matrix ensembles with orthogonal and symplectic symmetry [60]. In appendix C, we compute the Schwarzian and super Schwarzian path integral on a disc or trumpet; these computations are important inputs to the random matrix analysis of JT gravity and supergravity. In appendix D, we review Mirzakhani’s recursion relation for volumes of moduli spaces and generalize it to super Riemann surfaces. We arrive at the same recursion relation found in **section five**, thus confirming the relation between super JT gravity and the matrix model. In appendix E we make preliminary comments about non-perturbative effects in the matrix models, and in appendix F, we compare to results about the minimal string.

2. Random matrix ensembles and bulk topological theories

For the purposes of this paper, a matrix integral means an integral of the form

$$(2.1) \quad \int dM e^{-L \text{Tr} V(M)},$$

where M is an $L \times L$ matrix. Two types of data are needed to specify this integral.³ One is the potential function $V(M)$. In practice, it is more convenient to specify this function implicitly by giving the “spectral curve” or equivalently, the leading large L approximation to the density of eigenvalues. In our application, this piece of data will be determined by comparing to the gravity path integral on the disk topology, as we discuss in section 5 below.

The second piece of data is the symmetry class of matrices M over which we integrate. In random matrix theory, there are ten standard classes. The purpose of this section is to explain how the choice of symmetry class is related to a discrete choice of topological field theory that one can include in the bulk 2d gravity theory, and to explore some of the consequences.

We will start in section 2.1 by reviewing the symmetry classes of random matrix ensembles. In section 2.2 we then discuss the correspondence between topological field theories and random matrix classes for the simplest purely bosonic cases. In section 2.3 we discuss the strategy for the remaining more complicated cases. In section 2.4, we include $(-1)^F$ symmetry but no time-reversal. In section 2.5 we include both $(-1)^F$ symmetry and time-reversal, and in section 2.6 we include $\mathcal{N} = 1$ supersymmetry. Much of the complication in sections 2.4 and 2.5 has to do with demonstrating a type of reduction: correspondence with JT gravity in these cases follows from a correspondence in the three purely bosonic cases, together with spin and pin structure identities that we derive. No such reduction is possible in the supersymmetric cases of section 2.6.

2.1. Random matrix ensembles

2.1.1. The ensembles. There are 10 standard ensembles in random matrix theory, and each of them will play a role in the present paper. In each

³In our application, there is actually another type of discrete choice: in the application to JT gravity, we will regard the random matrix M as the Hamiltonian H of the boundary theory. In the application to JT supergravity, we will regard the matrix M as the supercharge Q such that $Q^2 = H$.

ensemble, one considers a class of $L \times L$ matrices M , with a symmetry group G that is either $U(L)$, $O(L)$, or $Sp(L)$, or a product of two groups of one of those types.

(i) In the original application of random matrix theory to nuclear physics, the random matrix M was interpreted as the Hamiltonian. In the absence of time-reversal symmetry,⁴ M is simply a random hermitian matrix M^i_j , $i, j = 1, \dots, L$, and the symmetry group that acts on the ensemble of such matrices is $G = U(L)$, acting by conjugation. If time-reversal symmetry \mathbb{T} is assumed, then $U(L)$ is reduced to the subgroup that commutes with \mathbb{T} . This is $O(L)$ if $\mathbb{T}^2 = 1$ and $Sp(L)$ if $\mathbb{T}^2 = -1$.⁵ Hermitian matrices that commute with \mathbb{T} correspond in the case of $O(L)$ to real symmetric matrices $M_{ij} = M_{ji}$. For $Sp(L)$, a Hamiltonian that commutes with \mathbb{T} takes the form $M^i_j = \varepsilon^{ik} M_{kj}$, where ε^{ik} is the invariant antisymmetric tensor of $Sp(L)$ and $M_{kj} = -M_{jk}$ is antisymmetric. Energy level statistics associated to these three classes are usually said to be of type GUE, GOE, or GSE (where U, O, or S stand for unitary, orthogonal, or symplectic, respectively, and GE stands for Gaussian ensemble, though we are interested in a generalization in which the measure is not really Gaussian).

These three ensembles are sometimes referred to as the Dyson ensembles [43]. Altland and Zirnbauer [44, 61] described seven more classes of random matrix ensemble:

(ii) In four cases, the symmetry group is a simple group and M is a second rank tensor of some kind. If $G = U(L)$, then M can be either a symmetric second rank tensor $M_{ij} = M_{ji}$ or an antisymmetric second rank tensor $M_{ij} = -M_{ji}$. Note that it does not matter if we consider a covariant or contravariant second rank tensor, since if M is of one type, then its adjoint M^\dagger is of the opposite type. A tensor M^i_j of mixed type is the hermitian matrix already considered in (i). If $G = O(L)$, M can be an antisymmetric tensor $M_{ij} = -M_{ji}$, and if $G = Sp(L)$, M can be a symmetric tensor $M_{ij} = M_{ji}$. Again, the tensors with the opposite symmetry properties were already considered in (i).

⁴Unitary symmetries (as opposed to the antiunitary symmetry of time-reversal) do not play an important role in classifying matrix ensembles. For example, if a \mathbb{Z}_2 symmetry g is assumed, one simply diagonalizes g and describes the Hamiltonian by one of the Wigner-Dyson ensembles in each eigenspace of g . We will encounter this situation shortly with $g = (-1)^F$.

⁵In nuclear physics, $\mathbb{T}^2 = 1$ for a nucleus with an even number of nucleons and $\mathbb{T}^2 = -1$ for an odd number.

(iii) Finally, there are three cases⁶ in which G is a product $U(L) \times U(L)$, $O(L) \times O(L)$, or $Sp(L) \times Sp(L)$, and M is a bifundamental M_{ij} , with one index transforming under the first factor and one under the second.⁷ These three examples can be generalized to add another integer ν , so that G is a product $U(L) \times U(L + \nu)$, $O(L) \times O(L + \nu)$, or $Sp(L) \times Sp(L + \nu)$. M is again taken to be a bifundamental.

2.1.2. The integration measure. For each of these random matrix ensembles, by the action of the symmetry group, it is possible to put M in a canonical form in terms of real “eigenvalues” λ_i . For some ensembles, the λ_i can vary independently, and for others, they have a two-fold or four-fold degeneracy. In the three original Wigner-Dyson ensembles, the independent λ_i are real-valued, but in the seven Altland-Zirnbauer ensembles, they can be chosen to be positive.

In each case, a G -invariant integral over M (by which one means an integral over each independent matrix element of M) reduces, after putting M in its canonical form, to an integral over the λ_i or more precisely over those λ_i that are independent. For the three Wigner-Dyson ensembles, the λ_i are simply the independent eigenvalues of a hermitian matrix (which may obey a time-reversal constraint, as explained earlier) and the measure for integration over the λ_i is of the form

$$(2.2) \quad dM \rightarrow \prod_{i < j} |\lambda_i - \lambda_j|^\beta \prod_k d\lambda_k$$

⁶These three ensembles had been discussed earlier in [45, 46] in studying the low energy spectrum of the massless Dirac operator in Euclidean signature coupled to a generic gauge field. Properties of this spectrum have implications for chiral symmetry breaking. In that application, the two factors of the symmetry group act on fermion modes of positive or negative chirality, respectively, and the random matrix is the chiral Dirac operator, which is a “bifundamental field” that exchanges the two types of mode. The symmetry group is $U(L) \times U(L)$, $O(L) \times O(L)$, or $Sp(L) \times Sp(L)$ depending on the nature of the fermion representation (complex, pseudoreal, or real).

⁷The reader might wonder why it is not possible to mix and match groups of different kinds, and to consider, for example, a bifundamental of $O(L) \times Sp(L)$. The problem is that the bifundamental of $O(L) \times Sp(L)$ is pseudoreal rather than real; as a real vector space its dimension is $2L^2$, not L^2 . By contrast, the dimension of the group $O(L) \times Sp(L)$ grows as L^2 for large L , not $2L^2$. So a bifundamental of $O(L) \times Sp(L)$ has of order L^2 real degrees of freedom that cannot be eliminated by the symmetry. By contrast, in the Wigner-Dyson and Altland-Zirnbauer ensembles, the number of independent degrees of freedom, modulo the symmetry, is always of order L .

with $\beta = 2, 1$, or 4 for GUE, GOE, or GSE. We will sometimes refer to these Dyson ensembles as β -ensembles. The β -dependent factor comes from the volume of the G orbits, as we describe presently. For the seven Altland-Zirnbauer ensembles, the measure is

$$(2.3) \quad dM \rightarrow \prod_{i < j} |\lambda_i^2 - \lambda_j^2|^\beta \prod_k |\lambda_k|^\alpha d\lambda_k,$$

with various pairs α, β . We will sometimes refer to these as (α, β) -ensembles.

Let us first consider the basic GUE ensemble. For $L = 1$, a hermitian matrix M is just a real number $M = \lambda$, and integration over M is the same as integration over λ . Now suppose that $L = 2$ and the canonical form of M is $\text{diag}(\lambda_1, \lambda_2)$. Such an M commutes with the group $G_0 = U(1) \times U(1)$ of diagonal matrices. Let \mathfrak{g}_0 be the Lie algebra of this subgroup, and let \mathfrak{g}_\perp be its orthocomplement, consisting of 2×2 hermitian matrices that are strictly off-diagonal. These matrices, when commuted with M , generate the tangent space to the orbit of M in the space of all hermitian matrices. Concretely, \mathfrak{g}_\perp is 2-dimensional, and the commutator of any $b \in \mathfrak{g}_\perp$ with M is proportional to $\lambda_1 - \lambda_2$. So the volume of the orbit is proportional to $(\lambda_1 - \lambda_2)^2$. More generally, for any L , let \mathfrak{g}_0 be the Lie algebra of diagonal matrices and \mathfrak{g}_\perp its orthocomplement, consisting of strictly off-diagonal matrices. Each pair of eigenvalues λ_i, λ_j is associated to a two-dimensional subspace of the full Hilbert space \mathcal{H} . Two generators of \mathfrak{g}_\perp act in each such subspace and their commutator with M is proportional to $\lambda_i - \lambda_j$. So the volume of the group orbit is a multiple of $\prod_{i < j} (\lambda_i - \lambda_j)^2$, leading to the measure (2.2) with $\beta = 2$.

The GOE and GSE cases are similar. For GOE, the canonical form of the real symmetric matrix M under the action of $G = O(L)$ is again $\text{diag}(\lambda_1, \dots, \lambda_L)$. G has one broken or off-diagonal generator for each eigenvalue pair λ_i, λ_j . The commutator of that generator with M is proportional to $\lambda_i - \lambda_j$, leading to the measure (2.2) with $\beta = 1$. For GSE, the eigenvalues are two-fold degenerate because of Kramers doubling of energy levels for $T^2 = -1$. So each independent variable λ_i over which one integrates actually represents a pair of eigenvalues. If the independent λ_i are all distinct, the subgroup of $G = \text{Sp}(L)$ that commutes with M is $G_0 = \text{Sp}(2)^{L/2}$. Let \mathfrak{g}_0 be the Lie algebra of G_0 and \mathfrak{g}_\perp its orthocomplement. If a pair λ_i, λ_j becomes equal, then a subgroup $\text{Sp}(2) \times \text{Sp}(2)$ of G_0 is enhanced to $\text{Sp}(4)$. As the difference of dimension between $\text{Sp}(4)$ and $\text{Sp}(2) \times \text{Sp}(2)$ is 4, for each pair $i \neq j$, there are 4 generators of \mathfrak{g}_0 whose commutator with M is proportional to $\lambda_i - \lambda_j$. This leads to the measure (2.3) with $\beta = 4$.

The seven Altland-Zirnbauer ensembles can be treated similarly, but there are two differences. First, for some of these ensembles there is symmetry enhancement when a single eigenvalue λ vanishes. If α is the increase in the dimension of the unbroken symmetry group when $\lambda \rightarrow 0$, then \mathfrak{g}_0 has that number of generators whose commutator with M is proportional to λ , and this leads to the factor $\prod_k \lambda_k^\alpha$ in the measure (2.3). Second, for the Altland-Zirnbauer ensembles, the group G can be used to flip separately the signs of the λ_i , which is why the λ_i can be chosen to be all positive. But this means that the same symmetry enhancement that occurs for $\lambda_i = \lambda_j$ must also occur for $\lambda_i = -\lambda_j$. If, therefore, β is the amount by which the dimension of the symmetry group is enhanced if $\lambda_i = \lambda_j$, the measure will contain a factor $(\lambda_i - \lambda_j)^\beta (\lambda_i + \lambda_j)^\beta = (\lambda_i^2 - \lambda_j^2)^\beta$. So in short α and β can be determined in all cases just by computing the dimensions of symmetry enhancements. (In all of these ensembles, symmetry enhancement occurs only if $\lambda_i = \pm\lambda_j$ or $\lambda_i = 0$.)

For illustration, we will compute α and β for one of the Altland-Zirnbauer ensembles – the case that $G = \mathrm{U}(L) \times \mathrm{U}(L)$ and M is an element of the bifundamental representation, which can be viewed as an $L \times L$ matrix with one $\mathrm{U}(L)$ factor acting on each side. (M corresponds to C in eqn. (2.63) below, not to the supercharge Q .) The canonical form of such a matrix under the action of $\mathrm{U}(L) \times \mathrm{U}(L)$ is $\mathrm{diag}(\lambda_1, \lambda_2, \dots, \lambda_L)$ where the λ_i can be assumed real and nonnegative. If the λ_i are all distinct, the unbroken symmetry is $G_0 = \mathrm{U}(1)^L$. If a single λ_i vanishes, an additional $\mathrm{U}(1)$ symmetry is restored, so $\alpha = 1$. If $\lambda_i = \lambda_j$ for some i, j , there is a symmetry enhancement from $\mathrm{U}(1) \times \mathrm{U}(1)$ to $\mathrm{U}(2)$, so $\beta = 2$. The other six Altland-Zirnbauer ensembles can be treated similarly. The results can be found in Tables 3 and 4 of section 2.6.2.

2.2. The purely bosonic cases

One can define three basic versions of bosonic JT gravity, and these are dual to matrix integrals of the three Dyson types: GUE, GOE, and GSE. The first version is JT gravity on orientable surfaces only, which is dual to a GUE-type matrix integral [35].

The remaining two Dyson ensembles (GOE and GSE) describe matrices that commute with a time-reversal symmetry. In AdS/CFT and related dualities, global symmetries of a boundary theory become gauge symmetries in the bulk description, with the restriction that a gauge transformation should act trivially on the boundary. For the case that the global symmetry is time-reversal T , the corresponding bulk theory should include T as a gauge

symmetry, which means that surfaces can be glued together with a reversal of orientation and unorientable manifolds are allowed. It is natural to guess that GOE and GSE ensembles should be related to bulk theories of this type.

Once the measure for summing over orientable manifolds has been fixed,⁸ there are two versions of the sum over possibly unorientable manifolds that are consistent with the general principles of topological field theory – these correspond to including or not including a factor of $(-1)^\chi$, where χ is the Euler characteristic. On a closed orientable manifold, χ is even and such a factor would have no effect. But on an unorientable manifold, χ can be odd and the factor is nontrivial.

The GOE and GSE matrix ensembles were shown to differ by precisely such a factor in [60], where χ labels the Euler characteristic of the 't Hooft double-line diagram in the perturbative expansion of the matrix integral. (For more on this point and its analog for Altland-Zirnbauer ensembles, see appendix B.) So it is natural to conjecture that GOE and GSE-like versions of the matrix integral studied in [35] will be dual to JT gravity on unorientable surfaces, with the $(-1)^\chi$ factor in the GSE case. Of course, these considerations are not limited to JT gravity. In the context of the $c < 1$ minimal string, the conjecture analogous to ours was made in [62–64]. In general the amplitudes for unoriented open strings with orthogonal or symplectic groups differ by a factor of $(-1)^\chi$.⁹

In studying the correspondence of matrix integrals and JT gravity, we will be interested in observables that correspond to path integrals over surfaces with boundaries. In this case there are two further points to consider. First, with an odd number of boundaries, χ would be an odd integer for an orientable surface, so in order to have a topological field theory that is trivial on orientable surfaces, it is convenient to replace the factor $(-1)^\chi$ by

⁸This involves an arbitrary parameter that controls the genus expansion, e^{-S_0} in random matrix theory or the string coupling constant in string theory.

⁹Unoriented open strings give orthogonal or symplectic symmetry depending on the sign of the operator Ω that exchanges the two ends of the string. Thus the two cases differ by the sign of $\text{Tr } \Omega \exp(-\beta H)$, which is the amplitude for a Mobius strip. On the other hand, the annulus partition function $\text{Tr } \exp(-\beta H)$ does not depend on the sign of Ω . Since the annulus and Mobius strip have χ differing by 1, a factor $(-1)^\chi$ in the path integral measure gives a relative sign between these two amplitudes. As we explain momentarily, on surfaces with boundary, it is useful to replace $(-1)^\chi$ with $(-1)^{n_c}$, where n_c is the number of crosscaps. Since a Mobius strip can be viewed as a disc with a crosscap attached, we have $(-1)^{n_c} = +1$ for an annulus and $(-1)^{n_c} = -1$ for a Mobius strip.

$(-1)^{n_c}$ where n_c is the number of crosscaps in the topological decomposition of the surface. (We will give an introduction to crosscaps below.) The factors $(-1)^X$ and $(-1)^{n_c}$ just differ by a minus sign for each boundary component; using $(-1)^{n_c}$ rather than $(-1)^X$ will let us avoid minus signs in the map between JT gravity and random matrix resolvents.

Second, in the sum over bulk geometries, the orientations of the boundaries should be regarded as gauge-invariant, because bulk gauge transformations are required to act trivially at the boundaries. So one can define gauge-invariant “orientation Wilson lines” that measure the change in orientation along a curve connecting the boundaries. These are \mathbb{Z}_2 -valued quantities, and with n boundaries there are $n - 1$ independent Wilson lines, so there are 2^{n-1} topologically distinct contributions. (The simplest case, with two boundaries, will be described in detail below, where we refer to the two possible geometries as the “double trumpet” and “twisted double trumpet.”) The implication is that even on an orientable manifold, the partition function of JT gravity with orientation-reversal gauged will differ from that of JT gravity without orientation-reversal gauged. It will be larger by a factor 2^{n-1} .

In later sections of this paper, we will compare the predictions of JT gravity on unorientable surfaces to GOE and GSE-like matrix integrals. But for the remainder of this section, we will focus on lining up other possible bulk theories with random matrix descriptions.

2.3. Strategy for the remaining cases

So far we have discussed the consequences of a symmetry \mathbb{T} . We would also like to explore the consequence of a symmetry $(-1)^F$ (the operator that assigns 1 to bosonic states and -1 to fermionic states), and eventually $\mathcal{N} = 1$ supersymmetry. In discussing the relevant matrix ensembles and their bulk duals, an essential role will be played by anomalies in the realization of the global symmetries $(-1)^F$ and \mathbb{T} . These anomalies determine both the topological field theory in the bulk and the random matrix theory class on the boundary.

In principle, the connection between random matrix classes and bulk topological field theories can be made abstractly, using these anomalies. However, we will find it convenient to use the SYK model as a concrete system that connects the concepts together. The SYK model is a quantum mechanical system of N Majorana fermions, and it will be useful for three reasons. First, although it is not a random matrix theory in the sense we mean it in this paper, it is “close enough,” in that the correlations of nearby

energy eigenvalues have a well-defined random matrix classification [65–68]. Second, at low energies it has an approximate bulk description that includes JT gravity or super JT gravity. And finally, it is flexible enough that its variants display all of the needed anomalies and exhaust the Dyson and Altland-Zirnbauer classification of random matrix theory.

The necessary anomalies have to do with the global symmetries $(-1)^F$ and time-reversal \mathbb{T} (if present), and they depend on the number of Majorana fermions N . If N is odd, there is an anomaly in a narrow sense – a violation of a classical symmetry – while if N is even, the expected symmetries are present, but they do not satisfy algebraic relations (such as $\mathbb{T}^2 = 1$) that would be expected classically.

The same anomalies are relevant in many different problems; examples without time-reversal symmetry include the Kitaev chain of Majorana fermions [69], the Ising model [70], and intersection theory on the moduli space of Riemann surfaces with boundary [71, 72]. The extension to include time-reversal also has various applications; for instance, see [73] for a time-reversal invariant version of the Kitaev chain with $\mathbb{T}^2 = 1$ and [74] for a version with $\mathbb{T}^2 = (-1)^F$. At a more abstract level, the mod 2 anomaly when $(-1)^F$ is the only symmetry and the mod 8 anomaly that arises when one includes time-reversal with $\mathbb{T}^2 = 1$ are related to the properties of Dirac operators in different dimensions and to the mod 2 periodicity of complex K-theory and the mod 8 periodicity of real K-theory [75]. In that context, the applications are too far-flung to be summarized here.

These symmetries and their anomalies have implications for the bulk theory. As described above, \mathbb{T} as a global symmetry of the boundary theory (and gauge symmetry of the bulk) means that one must sum over unorientable as well as orientable manifolds; $(-1)^F$ means that one must sum over spin structures. An anomaly in symmetries of a boundary theory must somehow be reflected in the couplings of the bulk dual. This in general happens as follows. Suppose that the boundary theory is formulated on a manifold X . In the bulk dual description, one sums over manifolds Y of one dimension higher. The theory on Y will then have couplings that are well-defined if Y has no boundary but that, if Y has a boundary X , are anomalous in a way that matches the anomaly of the original theory on X . For example, if the theory on X has continuous global symmetries that have 't Hooft anomalies (which would obstruct gauging those symmetries), then the bulk theory on Y has gauge fields with Chern-Simons couplings that are not gauge-invariant on a manifold with boundary. The analog of this for discrete symmetries – such as we consider here – is that the bulk description must include a topological field theory that captures the anomaly. This

is a topological field theory that is well-defined on a manifold Y without boundary, but anomalous when Y has a boundary.¹⁰

In what follows, we explain how this works for the discrete symmetries $(-1)^F$ and \mathbf{T} of the SYK model. We will then use the random matrix classification of SYK to relate bulk topological field theories to random matrix symmetry classes.

In discussing anomalies of $(-1)^F$ and \mathbf{T} without supersymmetry, we will encounter twelve cases, but we will not obtain essentially new random matrix classes. In the boundary theory, one finds combinations of the GOE, GUE, GSE ensembles [65–67]. In the bulk, we will have to analyze sums over spin and pin structures with weighting given by the topological field theory. However, the JT path integral does not depend on the spin structure, so we get a result that is simply the product of a bosonic JT path integral with a factor that comes from summing over spin or pin structures. By evaluating this factor, one reduces the duality to the three purely bosonic cases described above.

But in the ten distinct cases with $\mathcal{N} = 1$ supersymmetry, one finds new random matrix classes [66–68], which exhaust the full Altland-Zirnbauer classification. Moreover, JT supergravity has fermionic fields, so its path integral depends on the spin or pin structure. Hence, in the bulk JT supergravity, the sum over spin or pin structures does not just give a simple overall factor, but an essential and nontrivial part of the bulk theory.

2.4. Including $(-1)^F$ but not \mathbf{T}

2.4.1. The SYK model for even and odd N . To begin, we assume no symmetry except $(-1)^F$. Let us start with the SYK model with N Majorana fermions ψ_1, \dots, ψ_N . The action is

$$(2.4) \quad I = \int dt \left(\frac{i}{2} \sum_k \psi_k \frac{d\psi_k}{dt} - i^{q/2} \sum_{i_1 \dots i_q} j_{i_1 i_2 \dots i_q} \psi_{i_1} \psi_{i_2} \cdots \psi_{i_q} \right)$$

and the Hamiltonian is $H = i^{q/2} j_{i_1 i_2 \dots i_q} \psi_{i_1} \psi_{i_2} \cdots \psi_{i_q}$. Classically, there is, for all N , a symmetry $(-1)^F$ that acts by $\psi_k \rightarrow -\psi_k$. If q is not a multiple of 4, that is generically the only symmetry.

¹⁰The bulk factor $(-1)^{n_c}$ considered in section 2.2 is part of the definition of a purely bosonic topological field theory and could be treated in this framework, though it appears to be hard to do this in an enlightening way.

If N is even, then upon quantization we get a Clifford algebra of rank N . The Clifford algebra has an irreducible representation in a Hilbert space \mathcal{H} of dimension $2^{N/2}$. The symmetry operator $(-1)^F$ that anticommutes with the elementary fermions is a multiple of the product

$$(2.5) \quad \psi_1 \psi_2 \cdots \psi_N$$

of all N elementary fermion fields; this operator anticommutes with the elementary fermions. The SYK path integral,¹¹ on a circle of circumference β , computes the partition function $\text{Tr}_{\mathcal{H}} e^{-\beta H}$ if the fermions are antiperiodic in going around the circle, or $\text{Tr}_{\mathcal{H}} (-1)^F e^{-\beta H}$ if they are periodic. Antiperiodic or periodic fermions correspond to what we will call the Neveu-Schwarz (NS) or Ramond (R) spin structure. In general, both are nonzero when N is even.

A special case is that if $H = 0$, the path integral in the NS sector computes the dimension of Hilbert space:

$$(2.6) \quad \text{Tr}_{\mathcal{H}} 1 = 2^{N/2}.$$

Since this is the result with N Majorana fermions, the corresponding path integral with a single Majorana fermion equals $\sqrt{2}$.

If N is odd, the product in eqn. (2.5) commutes (rather than anticommutes) with the ψ_k , so it is a c -number in an irreducible representation of the algebra. The operation $\psi_k \rightarrow -\psi_k$ changes the sign of this c -number, so for odd N the Clifford algebra has two inequivalent irreducible representations, differing by $\psi_k \rightarrow -\psi_k$. Each of them has dimension $2^{(N-1)/2}$. Pick one of these representations and call it \mathcal{H} . It does not matter which one we pick since the Hamiltonian H , being an even function of the ψ_k , is invariant under $\psi_k \rightarrow -\psi_k$.

In the NS sector, the path integral of the SYK model on a circle of circumference β computes

$$(2.7) \quad \sqrt{2} \text{Tr}_{\mathcal{H}} e^{-\beta H}.$$

To see that the factor of $\sqrt{2}$ is necessary, consider the special case that $H = 0$. The path integral, with a factor $\sqrt{2}$ for each of N Majorana fermions, is then $2^{N/2}$. On the other hand, since the dimension of \mathcal{H} is $2^{(N-1)/2}$, we have $\text{Tr}_{\mathcal{H}} 1 = 2^{(N-1)/2}$, and therefore an extra factor of $\sqrt{2}$ is needed to match the SYK path integral. This factor means that for odd N , the SYK

¹¹We generally consider path integrals without operator insertions.

path integral does not have a natural interpretation as a trace in a Hilbert space. But it is still a well-defined path integral and it makes sense to ask what sort of bulk dual description it would have.

Now let us consider the Ramond sector. For odd N , the path integral of the SYK model in the Ramond sector (with no operator insertions) is actually identically zero. To see this, first note that if $H = 0$, so that the action (2.4) consists only of the kinetic energy, then each of the ψ_k has a zero-mode in the Ramond sector, so in all there are an odd number of zero-modes. Now including the Hamiltonian, a term that is proportional to H^r for some r has an insertion of qr fermions, which is an even number since q is even. An even number of these fermions can be paired up by propagators, and the remaining ones – also an even number – can be used to soak up zero-modes. Since we started with an odd number of fermion zero-modes, there is always an odd number left over. In particular, we can never soak up all of the zero-modes, and therefore the Ramond sector path integral of the odd N SYK model is identically zero.

By contrast, the Ramond sector path integral of the model with an insertion of an odd number of elementary fermion operators is generically nonzero, since with the help of such an insertion (along with the Hamiltonian) all of the zero-modes can be soaked up. Thus formally the Ramond sector path integral computes a generically nonzero expectation value $\langle \psi_k \rangle$ (where ψ_k is any one of the elementary fermion fields). This certainly qualifies as an anomaly, since it explicitly violates a hypothetical symmetry $(-1)^F$ that is supposed to act by $\psi_k \rightarrow -\psi_k$. This is related to the fact that for odd N , the $(-1)^F$ symmetry is lost at the quantum level, since $\psi_k \rightarrow -\psi_k$ exchanges two different representations of the Clifford algebra.

A summary of the above discussion is as follows. If we label the path integral of SYK with NS and R boundary conditions as $Z_{\text{NS}}(\beta)$ and $Z_{\text{R}}(\beta)$, then the path integrals are related to traces in the Hilbert space as

$$(2.8) \quad \begin{aligned} Z_{\text{NS}}(\beta) &= \begin{cases} \text{Tr } e^{-\beta H} & N \text{ even} \\ \sqrt{2} \text{Tr } e^{-\beta H} & N \text{ odd} \end{cases} \\ Z_{\text{R}}(\beta) &= \begin{cases} \text{Tr } (-1)^F e^{-\beta H} & N \text{ even} \\ 0 & N \text{ odd.} \end{cases} \end{aligned}$$

The trace is in a Hilbert space of dimension $2^{N/2}$ for even N , and dimension $2^{(N-1)/2}$ for odd N .

2.4.2. Initial random matrix considerations with $(-1)^F$ symmetry.

Let us now work out the relevant random matrix symmetry classes. For even

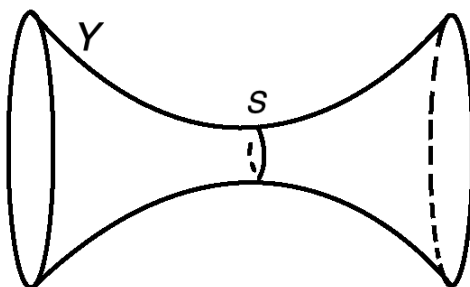


Figure 1: A double trumpet Y . At its “center” is a closed geodesic S .

N , we can choose a basis for the Hilbert space so that $(-1)^F = \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix}$. Then the only symmetry constraint is that H should commute with this matrix. A maximally random matrix consistent with the symmetry would be of type¹²

$$(2.9) \quad H = \begin{pmatrix} \text{GUE}_1 & 0 \\ 0 & \text{GUE}_2 \end{pmatrix}.$$

Here the subscripts on the two blocks indicate two independent GUE matrices. For odd N , there is no $(-1)^F$ symmetry, so the random matrix class is simply an unconstrained Hermitian matrix,

$$(2.10) \quad H = \text{GUE}.$$

We emphasize that although these random matrix ensembles have the correct symmetry properties, they do not describe the SYK model exactly, since SYK is not quite a random matrix theory. However, SYK was useful in motivating (2.8) (2.9) and (2.10), and we will see that random matrix ensembles of these types are precisely dual to appropriate refinements of JT gravity.

We will approach the relation to JT gravity in the following way. Correlation functions of observables $Z_{\text{NS}}(\beta)$ and $Z_{\text{R}}(\beta)$ in random matrix theory

¹²A subtlety here is that if $q = 2 \pmod 4$, then the SYK Hamiltonian actually anticommutes with a time-reversal operator. This affects the random matrix classification, but not at low energy (since a symmetry that anticommutes with H exchanges low and high energies), and it will not have a dual statement in JT gravity. However, to make the discussion above more precise for all energies, one can let H be a linear combination of $q = 0 \pmod 4$ and $q = 2 \pmod 4$ terms, so that H does not commute or anticommute with any version of T .

will be related to JT gravity path integrals on manifolds with boundaries with NS and R spin structures, and regularized lengths proportional to β . As a first case, consider the expectation value of the quantity $Z_{\text{NS}}(\beta)$. This is large in random matrix theory, in the sense that it is of order L , the rank of the matrix.¹³ The leading contribution to $\langle Z_{\text{NS}}(\beta) \rangle$ is dual to a JT gravity path integral on a spacetime with disc topology. But $Z_{\text{NS}}(\beta)$ also fluctuates in random matrix theory. The fluctuations can be measured by the connected two-point function

$$(2.11) \quad \langle Z_{\text{NS}}(\beta)^2 \rangle_c = \langle Z_{\text{NS}}(\beta)^2 \rangle - \langle Z_{\text{NS}}(\beta) \rangle^2.$$

In JT gravity, to compute this two-point function, we need a spacetime whose boundary consists of two circles, on each of which the holographic description would involve a path integral that generates a factor of $\text{Tr} \exp(-\beta H)$. But to get a contribution to the connected two-point function, the spacetime should be connected. The simplest possible topology is the “double trumpet,” topologically an annulus (fig. 1), and this makes the dominant contribution to the connected correlation function when L is large.

Now let us consider $Z_{\text{R}}(\beta)$. Its expectation value in JT gravity (without time-reversal symmetry) vanishes, since there is no oriented two-manifold whose boundary consists of a single circle with R spin structure. In random matrix theory, one interprets this as follows. For odd N , the R sector path integral is identically zero, as we have seen. For even N , this path integral is not identically zero, but it is zero on the average in random matrix theory: the random matrix class for H in (2.9) consists of two independent GUE matrices acting on the states of $(-1)^F = 1$ or $(-1)^F = -1$, and there is on average a cancellation between their contributions to $\text{Tr} (-1)^F e^{-\beta H}$.

Now let us consider in JT gravity the fluctuations of $Z_{\text{R}}(\beta)$ around its mean value of zero. These fluctuations can be measured by the two-point function $\langle Z_{\text{R}}(\beta)^2 \rangle$. This two-point function can receive contributions from connected oriented two-manifolds whose boundary consists of two circles and whose spin structure is of R type on each boundary. The simplest possible choice is again the double trumpet (with a different spin structure). The random matrix theory prediction is actually that for even N

$$(2.12) \quad \langle Z_{\text{R}}(\beta)^2 \rangle = \langle Z_{\text{NS}}(\beta)^2 \rangle_c \quad (N \text{ even}).$$

¹³The types of matrix integral that are actually dual to JT gravity are “double-scaled” matrix integrals, in which the analog of L is the parameter e^{S_0} that specifies the density of eigenvalues.

To compute in random matrix theory the product of traces on the right or left of eqn. (2.12), we have to sum over contributions of pairs of states which may have either eigenvalue 1 or -1 of $(-1)^F$. But pairs of states with opposite eigenvalues make no contribution to the connected correlation functions on either the left or the right of this equation, because in (2.9), the two blocks are considered statistically independent. On the other hand, pairs of states with the same eigenvalue of $(-1)^F$ make equal contributions on the left and right hand side of eqn. (2.12). For odd N , however, $\text{Tr} (-1)^F e^{-\beta H}$ is identically zero, so the expectation value of its square is also zero,

$$(2.13) \quad \langle Z_R(\beta)^2 \rangle = 0 \quad (N \text{ odd}).$$

By similar reasoning, for any N , the cross correlator

$$(2.14) \quad \langle Z_{NS}(\beta) Z_R(\beta) \rangle$$

vanishes. It vanishes in random matrix for even N because the states with $(-1)^F$ equal to 1 or -1 are statistically independent and contribute with opposite signs, and for odd N because $Z_R(\beta)$ is identically 0, without fluctuations. It vanishes in JT gravity because any oriented two-manifold with boundary has an even number of boundary components of R type.

Our initial goal will be to understand what topological refinement should be added to JT gravity to reproduce the predictions of eqns. (2.12) and (2.13).

2.4.3. Bulk description Of $(-1)^F$ symmetry and its anomaly. Let X be a compact 1-manifold (a circle or a disjoint union of circles) on which we want to study an SYK model or a related random matrix theory. X is endowed with a spin structure (of either NS or R type on each connected component of X) since the SYK model has fermions. In a bulk dual description (assuming no \mathbb{T} symmetry), we will sum in a JT-like model over oriented spin manifolds Y of boundary X , such that the spin structure of Y restricts on X to the spin structure of X . Given such a Y , we sum over equivalence classes of spin structures on Y , keeping fixed that of X . (The equivalence relation in this sum involves a subtlety that is explained later.) Such a sum represents the bulk dual of an SYK or random matrix calculation with a particular spin structure on X .

We want to find two versions of the sum over spin structures on X that will match the two cases of even and odd N . It will be useful to know the following. If Y is a two-dimensional compact oriented spin manifold without boundary, then one can consider the Dirac equation $\not{D}\lambda = 0$, for a spinor

field λ on Y of (say) positive chirality. Let ζ be the number of zero-modes of this equation, mod 2. It is a topological invariant, in the sense that it depends on the spin structure of Y but not on the choice of a Riemannian metric on Y . In the terminology of Atiyah and Singer, ζ is the mod 2 index¹⁴ of the Dirac operator on Y . For an introduction to the mod 2 index and in particular an explanation of why it is a topological invariant, see for instance section 3.2 of [76].

On an orientable two-manifold without boundary, there are two ways to sum over spin structures satisfying the general conditions of topological field theory. One can sum over spin structures assigning the same weight to each spin structure, or one can sum over spin structures with relative weights $(-1)^\zeta$. The reason that it makes sense to include a factor $(-1)^\zeta$ is that this factor is local in the relevant sense, though to understand that, one has to take a somewhat abstract view of what locality means. Usually in field theory, one has some fields Φ and some Lagrangian density $L(\Phi)$; the integrand of the path integral contains a factor $\Upsilon = \exp(-\int_Y d^Dx \sqrt{g} L(\Phi))$. This is local in the sense that if one varies Φ only in a small region $V \subset Y$, then Υ changes by a factor that only depends on what is happening inside V . Though this is far from obvious, $(-1)^\zeta$ is local in the same sense. The way we will use that is as follows. Suppose that we are given two spin structures on Y that coincide outside of V . Let ζ and ζ' be the mod 2 indices for these spin structures. Then the ratio $(-1)^\zeta/(-1)^{\zeta'}$ depends only on V (and the restriction of the two spin structures to V). It does not depend on anything that is happening outside V . This enables one to modify Y away from V in a way that makes computations simpler.

So on a two-manifold Y without boundary, a sum over spin structures with or without a factor of $(-1)^\zeta$ gives a topological field theory. However, our definition of ζ only made sense if Y has no boundary. (If Y has a boundary, then to make sense of the equation for a fermion zero-mode, one needs a boundary condition that mixes the two chiralities, so it is not possible to consider fermions of just one chirality.) If one tries to define $(-1)^\zeta$ on a two-manifold with boundary, then one runs into an anomaly. As is explained, for example, in section 3 of [72], this anomaly is precisely the anomaly of a system with an odd number of Majorana fermions on the boundary. Therefore, we expect that a simple sum over spin structures can be dual to an SYK-like model with even N , but that such a model with odd N requires including the factor $(-1)^\zeta$.

¹⁴An important point is that ζ is not the reduction mod 2 of an ordinary index or of any integer-valued topological invariant.

The only example in which we will actually need to calculate ζ is the following. Let Y be a two-torus, obtained by dividing the real xy plane by the identifications $x \cong x + 1$, $y \cong y + 1$. On Y , there are four spin structures, conveniently represented as (\pm, \pm) where the two signs appear in the periodicity or antiperiodicity relations

$$(2.15) \quad \lambda(x + 1) = \pm\lambda(x), \quad \lambda(y + 1) = \pm\lambda(y)$$

that a spinor field on Y should obey. Thus the $(+, +)$ spin structure means that λ is periodic in both directions, and in the other cases, λ is antiperiodic in one or both directions. Setting $z = x + iy$, the chiral Dirac equation is

$$(2.16) \quad \frac{\partial}{\partial \bar{z}} \lambda = 0.$$

In the case of the $(+, +)$ spin structure, there is a 1-dimensional space of solutions, with constant λ ; for the other cases, there are no solutions. So $\zeta = 1$ for the $(+, +)$ spin structure, and $\zeta = 0$ in the other cases.

We are now ready to analyze the contribution of the double trumpet in eqns. (2.12) and (2.13). Once one understands the double trumpet, it is not difficult to analyze any Y in a similar way. First of all, the spin structures on the two boundary components of the double trumpet are always of the same type. This is because the closed geodesic S at the center of the double trumpet (fig. 1) can be smoothly moved to the left boundary or the right boundary, so the spin structure is of the same type on each boundary as it is on S . More generally, it is true because in general, for any Y , the number of R type boundary components of Y is always even.

Now having fixed the spin structure on the two boundary components of Y , how many spin structures are there on Y , up to equivalence? This is a subtle question and we have to specify that we consider two spin structures on Y to be “equivalent” if they are equivalent under a redefinition of the fermion field that is trivial on the boundary of Y . The reason for the redefinition to be trivial on the boundary is that in general, in holographic duality, a “gauge equivalence” in a bulk description is supposed to involve a gauge transformation that is trivial along the boundary; the restriction of a gauge transformation to the boundary behaves as a global symmetry.

Given this, we can count the spin structures on Y that have a given restriction to the boundary as follows. Let Y_L and Y_R be “trumpets” (fig. 2) that can be glued together along “inner boundaries” S_L and S_R to build the “double trumpet” Y . The spin structure on Y_L or Y_R , once it is specified on

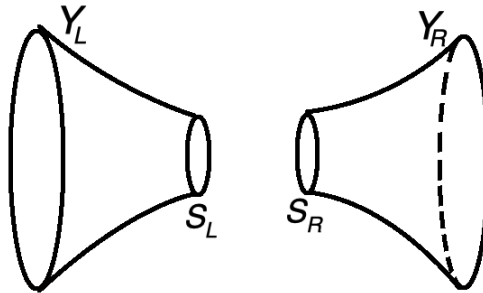


Figure 2: Two trumpets Y_L and Y_R . By gluing their “inner boundaries” S_L and S_R , one can build a “double trumpet.”

the outer boundary, is unique up to a unique equivalence,¹⁵ since any closed path in Y_L or in Y_R can be deformed to the outer boundary. However, when we glue S_L to S_R to build Y , a sign choice comes in; if λ_L and λ_R are fermion fields on S_L and S_R , we could identify λ_L with λ_R or with $-\lambda_R$. This sign can be eliminated if we are allowed to change the sign of the fermion field on all of Y_R (or Y_L), but this is not allowed since a gauge equivalence is supposed to be trivial on the outer boundaries.

Now we can explain eqns. (2.12) and (2.13) in the context of JT gravity. First of all, if N is even, then we are simply summing over spin structures on Y (restricted to agree with a given one on the boundary) without any signs. Each spin structure makes the same contribution, given by the double trumpet path integral of JT gravity, which is not sensitive to the spin structure. The spin structures that contribute to the left or right hand side of eqn. (2.12) are different, but there are 2 of them in each case. So the left or right hand side of eqn. (2.12) is equal to 2 times the double trumpet path integral of JT gravity.

For odd N , we have to deal with the factor of $(-1)^\zeta$. For this, first note that the two spin structures on Y (once the spin structures on the boundaries are given) are equivalent away from a small neighborhood of S , the geodesic at the center of the double trumpet (fig. 1). They differ by a minus sign that a fermion gets when it is parallel transported across S . Using the locality of $(-1)^\zeta$, as explained above, we can cut a small neighborhood of S out of the double trumpet and glue it into, for example, a two-torus (fig. 3). (Alternatively, and by a similar reasoning, one could glue together the two

¹⁵Here an equivalence is allowed to involve a field redefinition that is nontrivial along S_L or S_R , but not on the outer boundary.

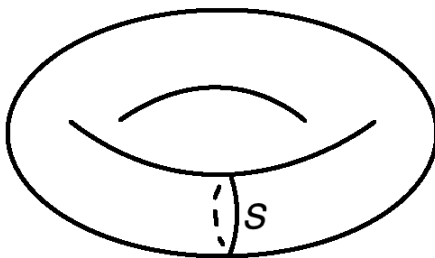


Figure 3: A small neighborhood of S has been cut out of the double trumpet and glued into a two-torus.

outer boundaries of the double trumpet, again arriving at a two-torus. By locality, that operation does not affect the comparison of the values of $(-1)^\zeta$ for a pair of spin structures that differ only near S .)

Now recall that spin structures on a two-torus are labeled as (\pm, \pm) , where here we can think of the first sign as the sign that a fermion gets when parallel transported around S , and the second sign as the sign that a fermion gets when it is parallel transported in a complementary direction around the torus, on a closed path that intersects S once. Moreover, $\zeta = 1$ for the $(+, +)$ spin structure and otherwise $\zeta = 0$. The first sign is $-$ or $+$ depending on whether the spin structures on the boundary components are of NS or R type (as noted above, the spin structure on S in the original double trumpet is of the same type as the spin structure on either of the two boundaries). Suppose that the boundary spin structure is of NS type. Then the spin structure on the torus after the regluing of fig. 3 is of type $(-, \pm)$, with some choice of the second sign. Regardless of the second sign, these spin structures have $\zeta = 0$ and so their contribution to the double trumpet path integral is insensitive to whether N is even or odd. Hence $\langle Z_{\text{NS}}(\beta)^2 \rangle$ is not sensitive to whether N is even or odd, in keeping with what one would expect from random matrix theory.

Suppose on the other hand that the boundary spin structures are of R type. Then the spin structures on the two-torus after regluing are of type $(+, \pm)$, with some choice of sign. One of them has $(-1)^\zeta = 1$ and one has $(-1)^\zeta = -1$, so their contributions cancel. Thus when N is odd, $\langle Z_{\text{R}}(\beta)^2 \rangle = 0$ in JT theory plus topological field theory, accounting for the claim in eqn. (2.13).

If one replaces the double trumpet with any connected two-manifold Y with two or more boundary components including some of R type, one can

make a similar argument, using for S any closed curve in the interior of Y that is homologous to a boundary of R type. Pairs of spin structures that differ by a minus sign that a fermion acquires when transported across S have opposite values of ζ and make canceling contributions. Thus for odd N , in JT gravity plus topological field theory, $\text{Tr}(-1)^F e^{-\beta H} = 0$, in the sense that all of its correlation functions vanish.

2.4.4. Reduction of cases with $(-1)^F$ symmetry to GUE. We would now like to consider a general correlation function, of the type

$$(2.17) \quad \langle Z_{\text{R}}(\beta_1) \dots Z_{\text{R}}(\beta_{n_{\text{R}}}) Z_{\text{NS}}(\beta_{n_{\text{R}}+1}) \dots Z_{\text{NS}}(\beta_n) \rangle_c.$$

It is somewhat simpler to work with a connected version of this correlator, by subtracting off all lower order correlation functions. As we will discuss in detail in section 4, in a matrix integral such connected correlations have a “genus” expansion

$$(2.18) \quad \langle Z_{\text{R}}(\beta_1) \dots Z_{\text{R}}(\beta_{n_{\text{R}}}) Z_{\text{NS}}(\beta_{n_{\text{R}}+1}) \dots Z_{\text{NS}}(\beta_n) \rangle_c \\ \simeq \sum_{g=0}^{\infty} \frac{Z_{g,n,n_{\text{R}}}(\beta_1, \dots, \beta_n)}{(e^{S_0})^{2g+n-2}}.$$

For an ordinary matrix integral, the expansion parameter would be L , but for the double-scaled matrix integrals that are relevant to JT gravity, it is a parameter that we refer to as e^{S_0} , proportional to the density of eigenvalues.

In random matrix theory, this expansion follows from 't Hooft's topological analysis [77] of matrix perturbation theory [78]. In JT gravity, this expansion corresponds to the sum over different topologies of connected manifold with n boundaries. The weighting by powers of e^{S_0} comes from the contribution of a term $-S_0\chi$ in the action of JT gravity. The “genus” g of a two-manifold is defined by $2 - 2g - n = \chi$, so that g is a non-negative integer in the case of an orientable surface, but in general may be a non-negative integer or half-integer.

For the ensembles defined by (2.8), (2.9), (2.10), we would like to relate the expansion coefficients $Z_{g,n,n_{\text{R}}}$ to similar expansion coefficients $Z_{g,n}$ in the ordinary GUE ensemble. One can do this as follows. First, in the case of even N , H is a direct sum of two independent matrices of GUE type, and connected correlation functions will be a sum of connected correlators for the two independent blocks. Including a factor of $(-1)^{n_{\text{R}}}$ for the block corresponding to fermionic states, one expects $Z_{g,n,n_{\text{R}}} = (1 + (-1)^{n_{\text{R}}})$ times the GUE answer. Similarly, in the case of odd N , (2.8) and (2.10) imply

that Z_{g,n,n_R} should be $(\sqrt{2})^n$ times the GUE answer if $n_R = 0$, and zero otherwise.

By rescaling the expansion coefficient e^{S_0} , we can multiply Z_{g,n,n_R} by a factor $(\text{const.})^{2g+n-2}$. It is convenient to choose the constant so that the “disk” amplitude $Z_{0,1,0}$ is equal to the standard GUE answer. This modifies the naive predictions described above to

$$(2.19) \quad Z_{g,n,n_R} = 2^{2g+n-2}(1 + (-1)^{n_R})Z_{g,n}^{\text{GUE}} \quad (N \text{ even})$$

$$(2.20) \quad Z_{g,n,n_R} = \begin{cases} 2^{g+n-1}Z_{g,n}^{\text{GUE}} & n_R = 0 \\ 0 & \text{else} \end{cases} \quad (N \text{ odd}).$$

This is the random matrix theory prediction. We would like to match it in the two variants of JT gravity that were described above, starting with the statement that JT gravity without spin structures matches $Z_{g,n}^{\text{GUE}}$ [35].

In fact, these relations follow from two identities for the sum over spin structures:

$$(2.21) \quad \sum_{\text{spin}} 1 = 2^{2g+n-2}(1 + (-1)^{n_R})$$

$$(2.22) \quad \sum_{\text{spin}} (-1)^\zeta = \begin{cases} 2^{g+n-1} & n_R = 0 \\ 0 & \text{else.} \end{cases}$$

In these expressions, the sum is over spin structures on an orientable surface with genus g and $n \geq 1$ boundaries, of which n_R are of R type and the rest are NS.

For $n = 0$, there are actually twice as many spin structures as these formulas would suggest, but one has to divide by a factor of 2 as a discrete Faddeev-Popov gauge fixing, because $(-1)^F$ has to be treated as a gauge symmetry. So for example, the formula 2^{2g+n-1} for the total number of spin structures is effectively still valid for $n = 0$. For $n > 0$, the overall $(-1)^F$ is not treated as a gauge symmetry because it is not gauged along the boundaries.

To derive these identities, one can start with fact that on a closed orientable surface, the number of even spin structures is $2^{g-1}(2^g + 1)$ and the number of odd ones is $2^{g-1}(2^g - 1)$. (These statements can be derived using the facts for the torus described above, and gluing together tori to form the genus g surface.) This implies that on a closed surface, $\sum 1 = 2^{2g}$ and $\sum (-1)^\zeta = 2^g$. To include the boundaries, we cut n holes in the surface and glue in n trumpets. The circles S_j along which we glue are treated as in the

discussion above. The sum over spin structure “orthogonal” to a given circle gives a factor of two in all cases except when the circle has R spin structure and we are summing with $(-1)^\zeta$. In this case we get zero. Note that the number of circles with R spin structure is always even.

To get the final results (2.21) and (2.22), one takes the product of the factor for the closed surface times all of the factors for the circles S_j associated to the n trumpets. A final factor of $\frac{1}{2}$ is needed due to the fact that this sum over spin structures counts as distinct situations that differ only by the sign of the spinor in the “interior” portion of the surface inside all of the circles S . This sign is pure gauge, so we divide by two to remove the overcounting. In the case of the double trumpet, the counting of spin structures, including this last factor, was explained in a slightly different way in section 2.4.3.

2.5. Including both $(-1)^F$ and T

2.5.1. Classical time-reversal symmetries in SYK-like models. The most obvious way to get time-reversal symmetry in the SYK model is to take q to be a multiple of 4. In this case, the classical SYK action (2.4) is invariant under a time-reversal symmetry that maps t to $-t$, and at $t = 0$ leaves the elementary fermions invariant,¹⁶

$$(2.23) \quad \mathsf{T}(\psi_k(0)) = \psi_k(0).$$

Classically, the time-reversal symmetry defined this way satisfies some obvious relations

$$(2.24) \quad \mathsf{T}^2 = 1, \quad \mathsf{T}(-1)^F = (-1)^F \mathsf{T},$$

where $(-1)^F$ is understood to satisfy

$$(2.25) \quad \left((-1)^F \right)^2 = 1.$$

As we will discuss, there are anomalies in these statements at the quantum level.

If instead q is congruent to 2 mod 4, then there is no time-reversal symmetry that acts precisely as in (2.23). However, even for q of this form, it

¹⁶Classically, we view T as a transformation of observables and write its action on an observable ψ as $\mathsf{T}(\psi)$. Quantum mechanically, we will view T as a linear operator acting on quantum states, and write the action on operators as $\psi \rightarrow \mathsf{T}\psi\mathsf{T}^{-1}$.

is possible to constrain the SYK couplings to respect such a symmetry. There are actually two ways to do this. In one approach, one divides the two sets of fermions into two subsets, say ψ_i , $i = 1, \dots, N_+$ and $\tilde{\psi}_j$, $j = 1, \dots, N_-$, with $N_+ + N_- = N$, the total number of Majorana fermions. Then one defines \mathbb{T} to act at $t = 0$ as

$$(2.26) \quad \mathbb{T}(\psi_i(0)) = \psi_i(0), \quad \mathbb{T}(\tilde{\psi}_j(0)) = -\tilde{\psi}_j(0).$$

This is not a symmetry of the generic SYK action (2.4). However, we can constrain the SYK coupling parameters $j_{i_1 i_2 \dots i_q}$ to respect the symmetry of eqn. (2.26) and otherwise to be random variables with specified variance. It is reasonable to expect that such a model can have SYK-like behavior, in the sense that the long time behavior is dominated by a Schwarzian mode, as in the case of SYK.

This type of time-reversal symmetry obeys at the classical level the same algebraic relations (2.24) and (2.25) as before. It turns out to have the same anomalies as in the previous case, with N replaced by $N_+ - N_-$. So from a topological field theory point of view, this generalization does not seem to add anything new, but in section 2.6, we will see that it is useful to be familiar with it.

Still with q congruent to 2 mod 4, it is actually also possible to impose an essentially different sort of time-reversal symmetry.¹⁷ For this, we assume N to be even, and we divide the fermions into two groups of equal size, say $\psi_1, \dots, \psi_{N/2}$ and $\tilde{\psi}_1, \dots, \tilde{\psi}_{N/2}$. Then we take \mathbb{T} to act by

$$(2.27) \quad \mathbb{T}(\psi_k(0)) = \tilde{\psi}_k(0), \quad \mathbb{T}(\tilde{\psi}_k(0)) = -\psi_k(0).$$

Now we have not $\mathbb{T}^2 = 1$ at the classical level, but

$$(2.28) \quad \mathbb{T}^2 = (-1)^F.$$

Again, the usual SYK model does not respect such a symmetry, but if we choose the coupling parameters to respect this symmetry (and otherwise to be random variables of specified variance), it is reasonable to expect to find SYK-like behavior.

¹⁷We could do the same for q a multiple of 4, but the import would be different. The SYK model with q a multiple of 4 unavoidably has a standard \mathbb{T} symmetry with $\mathbb{T}^2 = 1$, so any other \mathbb{T} symmetry would be the standard one times some global symmetry.

As remarked in section 2.3, when a boundary theory has time-reversal symmetry, a bulk dual description will involve a sum over possibly unoriented two-manifolds. There are two essentially different generalizations of the notion of a spin structure to unorientable manifolds. In Euclidean signature (where we will work in our study of bulk duals), these are called pin^+ structures and pin^- structures, and they correspond respectively to theories that have time-reversal symmetries that at the classical level satisfy $\mathbb{T}^2 = (-1)^F$ or $\mathbb{T}^2 = 1$, respectively.

Thus the usual SYK model with q a multiple of 4 may be expected to have a bulk dual that involves a sum over pin^- structures. Instead, an SYK-like model with $\mathbb{T}^2 = (-1)^F$ will potentially have a bulk dual that involves a sum over pin^+ structures.

With $\mathbb{T}^2 = 1$, the anomalies depend on the value of N (or more generally $N_+ - N_-$) mod 8, so there are 8 different cases. This matches the fact that there are 8 topological field theories based on a sum over pin^- structures in two dimensions. With $\mathbb{T}^2 = (-1)^F$, N has to be even and it turns out that only the value of N mod 4 matters, so there are two cases, with N congruent to 0 or 2 mod 4. This matches the fact that two topological field theories can be made by summing over pin^+ structures. We will describe the anomalies and the topological field theories, and try to match what we learn with what one would expect from random matrix theory.

2.5.2. Anomalies when $\mathbb{T}^2 = 1$ at the classical level. We will consider a theory – such as the SYK model with q a multiple of 4 – that has a symmetry group $\mathbb{Z}_2 \times \mathbb{Z}_2$ at the classical level. One \mathbb{Z}_2 is generated by $(-1)^F$, and the other by a time-reversal transformation \mathbb{T} that commutes with the elementary fermions and obeys $\mathbb{T}^2 = 1$. The other nontrivial element of the group is $\mathbb{T}' = \mathbb{T}(-1)^F$, which anticommutes with elementary fermions and satisfies

$$(2.29) \quad (\mathbb{T}')^2 = 1, \quad (-1)^F = \mathbb{T}\mathbb{T}'.$$

We write $\mathbb{Z}_2^{\mathbb{T}} \times \mathbb{Z}_2^F$ for the group generated by \mathbb{T} and $(-1)^F$.

It will be convenient to work with the fermion fields $\chi_k = \sqrt{2}\psi_k$, which obey a conventionally normalized Clifford algebra

$$(2.30) \quad \{\chi_k, \chi_l\} = 2\delta_{kl}.$$

This will minimize factors of $\sqrt{2}$ in the following formulas. Also, it is convenient to use the conventional 2×2 Pauli matrices $\sigma_1, \sigma_2, \sigma_3$, where σ_2 is imaginary and antisymmetric and σ_1, σ_3 are real and symmetric. We will

now work out the anomalies in $(-1)^F$ and T by working out their operator representations in terms of the fundamental SYK fermions.

The case that N is even. For $N = 2$, we can represent the Clifford algebra of the χ 's by 2×2 matrices

$$(2.31) \quad \chi_1 = \sigma_1, \chi_2 = \sigma_2.$$

For $N = 4$, we need 4×4 matrices. We can think of these as matrices that act on a pair of qubits, where χ_1, χ_2 act on the first qubit as before and χ_3, χ_4 are new:

$$(2.32) \quad \chi_1 = \sigma_1 \otimes 1, \chi_2 = \sigma_2 \otimes 1, \chi_3 = \sigma_3 \otimes \sigma_1, \chi_4 = \sigma_3 \otimes \sigma_2.$$

Every time we increase N by 2, we add another qubit, replace the existing χ_k by $\chi_k \otimes 1$, and add two new χ 's

$$(2.33) \quad \chi_{N-1} = \sigma_3 \otimes \sigma_3 \otimes \cdots \otimes \sigma_3 \otimes \sigma_1, \quad \chi_N = \sigma_3 \otimes \sigma_3 \otimes \cdots \otimes \sigma_3 \otimes \sigma_2.$$

The purpose of the factors of σ_3 is to make sure that the new χ 's anti-commute with the previous ones. The χ_k are real for odd k and imaginary for even k . These formulas give an irreducible representation of the even N Clifford algebra in a Hilbert space of dimension $2^{N/2}$.

An operator $(-1)^F$ that anticommutes with all of the χ_k and satisfies $((-1)^F)^2 = 1$ is

$$(2.34) \quad (-1)^F = i^{N(N-1)/2} \chi_1 \chi_2 \cdots \chi_N.$$

Letting K denote complex conjugation, an antiunitary time-reversal transformation T that commutes with the χ_k and therefore (if q is a multiple of 4) with the usual SYK Hamiltonian is

$$(2.35) \quad T = \begin{cases} K \chi_2 \chi_4 \chi_6 \cdots \chi_N & N = 0 \pmod{4} \\ K \chi_1 \chi_3 \chi_5 \cdots \chi_{N-1} & N = 2 \pmod{4}. \end{cases}$$

The other symmetry is just $T' = T(-1)^F$. Based on these formulas, we find that

$$(2.36) \quad T^2 = \begin{cases} 1, & \text{if } N = 0, 2 \pmod{8} \\ -1, & \text{if } N = 4, 6 \pmod{8} \end{cases}$$

and

$$(2.37) \quad \mathbb{T}(-1)^F = \begin{cases} (-1)^F \mathbb{T}, & \text{if } N = 0, 4 \pmod{8} \\ -(-1)^F \mathbb{T}, & \text{if } N = 2, 6 \pmod{8}. \end{cases}$$

Thus, for even N , there are anomalies in the statements $\mathbb{T}^2 = 1$, $\mathbb{T}(-1)^F = (-1)^F \mathbb{T}$, unless N is a multiple of 8, and these anomalies depend nontrivially on $N \pmod{8}$.

We now discuss the random matrix classes associated with these symmetries. There is a unitary symmetry $(-1)^F$, and we can treat the Hamiltonian as a random matrix in each block labeled by the eigenvalue of $(-1)^F$. If $N = 0$ or $4 \pmod{8}$, then \mathbb{T} commutes with $(-1)^F$ and constrains each block separately. In more detail, for $N = 0 \pmod{8}$, $\mathbb{T}^2 = 1$ so the statistics in each block are GOE-like. For $N = 4 \pmod{8}$, $\mathbb{T}^2 = -1$ and the statistics in each block are GSE-like. On the other hand, for $N = 2$ or $6 \pmod{8}$, \mathbb{T} anticommutes with $(-1)^F$ and exchanges the two blocks, so each block has the same energy levels. However, in these cases, \mathbb{T} does not constrain the Hamiltonian within any one block, and the statistics in either block are GUE-like.

The case that N is odd. The story is similar for odd N , though some details are different. As we noted in section 2.4.1, $(-1)^F$ does not act in an irreducible representation. Since $(-1)^F = \mathbb{T}\mathbb{T}'$, it will not happen that both \mathbb{T} and \mathbb{T}' do act. What happens is that for each odd N , one of them acts within an irreducible representation, and its square is again ± 1 .

To represent the Clifford algebra, we can represent $\chi_1, \dots, \chi_{N-1}$ as before in a Hilbert space of dimension $2^{(N-1)/2}$. Then we define

$$(2.38) \quad \chi_N = \pm i^{(N-1)/2} \chi_1 \chi_2 \cdots \chi_{N-1},$$

which is always real. Either choice of sign gives a representation of the rank N Clifford algebra in a Hilbert space of dimension $2^{(N-1)/2}$. These representations are irreducible, and they are inequivalent since they have opposite values of the central element

$$(2.39) \quad \chi_1 \chi_2 \cdots \chi_N$$

of the Clifford algebra. We choose one of them; nothing that follows depends on this choice.

We define a time-reversal operator $\widehat{\mathbb{T}}$ as

$$(2.40) \quad \widehat{\mathbb{T}} = \mathbb{K}\chi_1\chi_3\chi_5 \cdots \chi_N.$$

It obeys

$$(2.41) \quad \widehat{\mathbb{T}}\chi_k\widehat{\mathbb{T}}^{-1} = \begin{cases} \chi_k & \text{if } N = 1, 5 \pmod 8 \\ -\chi_k & \text{if } N = 3, 7 \pmod 8. \end{cases}$$

and

$$(2.42) \quad \widehat{\mathbb{T}}^2 = \begin{cases} 1 & \text{if } N = 1, 7 \pmod 8 \\ -1 & \text{if } N = 3, 5 \pmod 8. \end{cases}$$

Eqn. (2.41) shows that $\widehat{\mathbb{T}} = \mathbb{T}$ if $N = 1, 5 \pmod 8$, while $\widehat{\mathbb{T}} = \mathbb{T}'$ if $N = 3, 7 \pmod 8$. This and the sign of $\widehat{\mathbb{T}}^2$ distinguishes the four cases mod 8.

The random matrix classes are somewhat simpler in this case. Since there is no nontrivial $(-1)^F$ symmetry, there is only one block. For $N = 1$ or $7 \pmod 8$, $\mathbb{T}^2 = 1$, so the symmetry group that commutes with \mathbb{T} is $O(L)$, and the statistics are of GOE type. For $N = 3$ or $5 \pmod 8$, $\mathbb{T}^2 = -1$, and the statistics are of GSE type. This classification, together with the even N case, is summarized in table 1 of section 2.5.3.

Further comments. We have described anomalies that depend on the value of $N \pmod 8$, but the reader might ask if there could be a more subtle anomaly that we have not found that remains even if N is a multiple of 8. The most powerful way to show that this is not the case is the following [79]. If χ_α , $\alpha = 1, 2, \dots, 8$ is a group of 8 Majorana fermions, all commuting with \mathbb{T} , then a generic quartic SYK-like Hamiltonian¹⁸ $\Delta H = \sum_{\alpha\beta\gamma\delta} t_{\alpha\beta\gamma\delta} \chi_\alpha\chi_\beta\chi_\gamma\chi_\delta$ has a unique ground state, which moreover (if the coefficients $t_{\alpha\beta\gamma\delta}$ obey a suitable inequality) is invariant under both \mathbb{T} and $(-1)^F$. That means that, by taking the coefficients in ΔH to be large enough, these 8 Majorana fermions can be removed from an effective low energy description of the system, without breaking any symmetries. Since anomalies in global symmetries can always be understood in terms of any description of the system that is valid at low energies, this implies that anomalies in the symmetries of this system can only depend on the value of $N \pmod 8$. (For N a multiple of 8, it is possible to pick a different representation of the Clifford algebra in which the χ_k are all real and therefore time-reversal can act more simply as $\mathbb{T} = \mathbb{K}$.)

¹⁸A simple one that does the job is $\Delta H = -\chi_1\chi_2\chi_3\chi_4 - \chi_5\chi_6\chi_7\chi_8 - \chi_1\chi_2\chi_5\chi_6 - \chi_1\chi_3\chi_5\chi_7$.

Still assuming that the symmetry group is $\mathbb{Z}_2^{\text{T}} \times \mathbb{Z}_2^{\text{F}}$, we now want to consider the case that at the classical level there are Majorana fermions $\chi_1, \dots, \chi_{N_+}$ with

$$(2.43) \quad \text{T}\chi_k\text{T}^{-1} = \chi_k,$$

and additional Majorana fermions $\tilde{\chi}_1, \dots, \tilde{\chi}_{N_-}$ with

$$(2.44) \quad \text{T}\tilde{\chi}_k\text{T}^{-1} = -\tilde{\chi}_k.$$

This generalization can be analyzed without much effort. First of all, there is no anomaly if $N_+ = N_- = 1$. We can work in a two-dimensional Hilbert space with $\chi = \sigma_1$, $\tilde{\chi} = \sigma_2$, along with $\text{T} = \text{K}$, $(-1)^{\text{F}} = \sigma_3$. All expected commutation relations are satisfied, and in particular $\text{T}^2 = 1$ and $\text{T}(-1)^{\text{F}} = (-1)^{\text{F}}\text{T}$.

More generally, if we have any system of N Majorana fermions with a time-reversal symmetry T , and two of them, say χ and $\tilde{\chi}$, transform under T with opposite signs, then a perturbation to the Hamiltonian $\Delta H = im\chi\tilde{\chi}$ is T -invariant, and is invariant under $(-1)^{\text{F}}$ if N is even (so that such a symmetry exists). For large m , χ and $\tilde{\chi}$ are removed from the low energy description, without breaking any symmetry. Such an operation cannot affect any anomalies, and it reduces N_+ and N_- by 1. So anomalies can only depend on the difference $N_+ - N_-$. Combining this with the previous argument, we see that anomalies depend precisely on the value of $N_+ - N_- \pmod 8$.

There is one last important comment. A low energy observer who does not have access to microscopic fermion fields will not be able to distinguish an anomaly of N from an anomaly of $8 - N$. This means, in particular, that JT gravity together with topological field theory are not sensitive to the sign of the anomaly.

Exchanging N with $8 - N$ changes the sign of the anomaly, but one can compensate for this by exchanging T with T' , which also changes the sign of the anomaly, as it exchanges (N_+, N_-) with (N_-, N_+) . For even N , the low energy observer sees both T and T' symmetry, but – without access to a microscopic description – has no way to know which is which. For odd N , the low energy observer has access to only one of T and T' , and has no way to know which it is. Whether N is even or odd, one cannot distinguish N from $8 - N$ without access to the elementary fermions.

Note that along with elementary fermion fields χ_k , whose transformation under time-reversal we have discussed, SYK-like models have hermitian

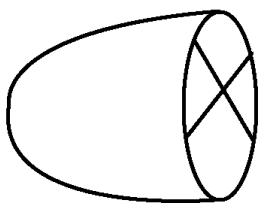


Figure 4: This picture is meant to symbolically convey the idea of building \mathbb{RP}^2 by closing off a disc with a “crosscap.”

fermion fields such as $i\chi_k\chi_l\chi_m$ which transform under time-reversal with an opposite sign. To distinguish \mathbb{T} from \mathbb{T}' , so that one can distinguish N from $8 - N \bmod 8$, one needs access to a fermion field whose time-reversal properties are related to those of the elementary fermions in a known way. An example in which this is possible in a macroscopic description is the supersymmetric SYK model. A low energy observer has access to the fermion in the super-Schwarzian multiplet, whose time-reversal properties are related in a simple way to those of the elementary fermions. The super-Schwarzian multiplet is part of the description by JT supergravity. Consequently, JT supergravity plus topological field theory can distinguish N from $8 - N \bmod 8$, as we will see in detail.

2.5.3. Topological field theory interpretation for cases with $\mathbb{T}^2 = 1$ classically. To define fermions on an unorientable manifold, one needs a generalization of a spin structure that is called a pin^+ structure or a pin^- structure. The meaning of pin^+ or pin^- is that if \mathbb{R} is a spatial reflection, then \mathbb{R}^2 commutes or anticommutes with fermion fields. A relativistic theory has a \mathbb{T} symmetry if and only if it has an \mathbb{R} symmetry, but there is a perhaps surprising sign reversal: if \mathbb{T}^2 commutes with elementary fermions, then \mathbb{R}^2 anticommutes with them, and vice-versa.¹⁹ So $\mathbb{T}^2 = 1$ in Lorentz signature is associated after Wick rotation to pin^- structures, while $\mathbb{T}^2 = (-1)^F$ is associated to pin^+ structures. For a detailed introduction to pin^- and pin^+ structures, see appendix A of [76].

¹⁹ The usual explanation in terms of the Dirac equation for a fermion field ψ is as follows. One starts with a Clifford algebra $\{\gamma_\mu, \gamma_\nu\} = \pm 2\eta_{\mu\nu}$, where the choice of sign will lead to pin^+ or pin^- . Suppressing all space coordinates except one, a relativistic theory will have a symmetry $\mathbb{T} : \psi(t, x) \rightarrow \gamma_0\psi(-t, x)$ if and only if it has a symmetry $\mathbb{R} : \psi(t, x) \rightarrow \gamma_1\psi(t, -x)$. Since $\gamma_0^2 = -\gamma_1^2$, one of these operators squares to 1 if and only if the other squares to $(-1)^F$.

The topological invariant for a compact two-dimensional pin^- manifold Y without boundary that generalizes $(-1)^\zeta$ for spin manifolds is $\exp(-i\pi\eta/2)$, where η is the Atiyah-Patodi-Singer eta-invariant of the self-adjoint operator $i\not{D}$. For any Y , $\exp(-i\pi\eta/2)$ is an eighth root of unity.²⁰ A basic example for which η has the minimum value²¹ $\pm 1/2$ is \mathbb{RP}^2 .

There are many possible descriptions of \mathbb{RP}^2 , but the one that is most useful for us is as follows. Begin with a closed disc D , whose boundary is a circle S , say parametrized by an angle θ , with $\theta \cong \theta + 2\pi$. One can build from D a two-manifold without boundary by identifying antipodal points on S , that is by making the further identification $\theta \cong \theta + \pi$. This is often described by saying that D is closed off with a “crosscap.” As the result is difficult to draw convincingly, it is often depicted symbolically as in fig. 4. The resulting compact manifold is called \mathbb{RP}^2 (real projective two-space). It is unorientable, with Euler characteristic 1, so its genus, defined by $\chi = 2 - 2g$, is $g = 1/2$. By “gluing in a crosscap” to a two-manifold Y , one means that one cuts out of Y an open ball and closes it off with a crosscap in the sense just described. Topologically, any compact unorientable two-manifold is an oriented one with a certain number of crosscaps glued in (in fact, one or two crosscaps is enough). Every time one glues in a crosscap, the Euler characteristic is reduced by 1. So the surfaces with odd Euler characteristic are the ones with an odd number of crosscaps glued in; they are the ones with half-integer g .

Since $\exp(-i\pi\eta/2)$ is an eighth root of unity, and has the appropriate locality properties, one can build 8 different topological field theories on a pin^- manifold Y without boundary by summing over pin^- structures on Y

²⁰This can be proved along lines explained in detail in appendix C of [76] for an analogous question in four dimensions. The two-dimensional case is slightly simpler. In brief, we want to prove that, for any pin^- bundle $\mathcal{P} \rightarrow Y$, $(\exp(-i\pi\eta/2))^8 = 1$, or equivalently that η is a multiple of $1/2$. Let ε be the orientation bundle of Y (a real line bundle with holonomy -1 around any orientation-reversing loop) and let $\mathcal{P}' = \mathcal{P} \otimes \varepsilon$ be the complementary pin^- bundle to \mathcal{P} . Then $\eta(\mathcal{P}) + \eta(\mathcal{P}') = 0 \pmod 4$, because of general properties of the spectrum of the Dirac operator, but also $4\eta(\mathcal{P}) = 4\eta(\mathcal{P}') \pmod 4$, because the real vector bundle $\varepsilon^{\oplus 4} \rightarrow Y$ is trivial for any two-manifold Y . These steps are explained more fully in [76].

²¹Consider an orbifold T^3/\mathbb{Z}_2 , where T^3 is a three-torus and \mathbb{Z}_2 acts as -1 on each of the three coordinates. The \mathbb{Z}_2 action has eight fixed points. Removing a small open neighborhood of each, one gets a compact three-manifold W whose boundary consists of 8 copies of \mathbb{RP}^2 . Applying the Atiyah-Patodi-Singer index theorem to W , one learns that $\eta(\mathbb{RP}^2) = \pm 1/2$, with a sign that depends on the choice of pin^- structure. For the details of this argument (in an analogous four-dimensional case), see appendix C of [76].

with a factor of $\exp(-i\pi N\eta/2)$. When Y has a boundary, $\exp(-i\pi N\eta/2)$ has an anomaly that precisely matches the anomaly of N Majorana fermions on the boundary of Y . This is schematically explained in section 5 of [76]. For a more abstract explanation based on cobordism invariance, see [80].

So we expect that in a bulk description of the SYK model with the usual sort of time-reversal symmetry, we should include a sum over pin^- structures with a factor of $\exp(-i\pi N\eta/2)$. We will call the sum over the pin^- structures on a manifold, weighted by $\exp(-i\pi N\eta/2)$, the “ pin^- sum.” We will denote this sum on a manifold Y as $F_Y(N)$. If Y is orientable, then a pin^- structure on Y is the same as a spin structure, and $\exp(-i\pi\eta/2)$ reduces to $(-1)^\zeta$. So for orientable Y , the pin^- sum is just the sum over spin structures weighted by $(-1)^{N\zeta}$. This is the same sum that we studied in sections 2.4.3 and 2.4.4, and is only sensitive to $N \bmod 2$. For unorientable Y , the pin^- sum can detect the value of $N \bmod 8$.

However, as in section 2.5.2, JT gravity and topological field theory cannot distinguish N from $8 - N \bmod 8$. Here we use the fact that pin^- bundles over Y come in pairs, in the following way. If \mathcal{P} is a pin^- bundle over Y , then there is another pin^- bundle \mathcal{P}' over Y , with the property that parallel transport of a fermion around any closed loop $\gamma \subset Y$ gives the same result up to sign. The sign is $+1$ or -1 depending on whether the orientation of Y is invariant or is reversed in going around γ . The abstract way to say this is that $\mathcal{P}' = \mathcal{P} \otimes \varepsilon$, where ε is the orientation bundle of Y . Then $\eta(\mathcal{P}) = -\eta(\mathcal{P}') \bmod 4$, so $\exp(-\pi i N \eta(\mathcal{P})/2) = \exp(-\pi i (8 - N) \eta(\mathcal{P}')/2)$. In other words, exchanging \mathcal{P} with \mathcal{P}' has the same effect as replacing N by $8 - N \bmod 8$. So once one sums over pin^- structures, the distinction between N and $8 - N$ disappears. Moreover, it disappears by the same mechanism as in section 2.5.2, since exchanging \mathcal{P} with \mathcal{P}' amounts to exchanging fermions that transform with opposite signs under \mathbb{T} .

For a variety of reasons, an important special case is the pin^- sum of \mathbb{RP}^2 . Since \mathbb{RP}^2 has two pin^- structures with $\eta = \pm 1/2$, its pin^- sum is $F_{\mathbb{RP}^2}(N) = 2 \cos(2\pi N/8)$. In particular, $F_{\mathbb{RP}^2}(2) = 0$ and $F_{\mathbb{RP}^2}(N + 4) = -F_{\mathbb{RP}^2}(N)$ for any N .

These facts have several interesting implications. Any two-manifold Y can be built from an oriented one Y_0 by gluing in some number n_c of cross-caps. Here n_c is not uniquely determined, but it is uniquely determined mod 2; in fact, n_c is congruent mod 2 to the Euler characteristic $\chi(Y)$. Using the locality properties of $\exp(-i\pi\eta/2)$, one can show that

$$(2.45) \quad F_Y(N) = F_{Y_0}(N) F_{\mathbb{RP}^2}(N)^{n_c}.$$

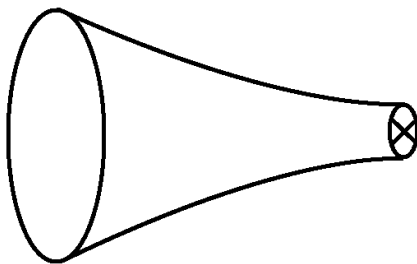


Figure 5: A trumpet that ends on a crosscap; we will call this the crosscap spacetime.

Since $F_{Y_0}(N)$ only depends on $N \bmod 2$, and $F_{\mathbb{R}P^2}(N+4) = -F_{\mathbb{R}P^2}(N)$, we get

$$(2.46) \quad F_Y(N+4) = (-1)^{n_c} F_Y(N).$$

In other words, shifting N by 4 simply gives a factor of -1 for every crosscap. The random matrix counterpart of this is discussed in appendix B.

Another interesting consequence is the following. Since Y is unorientable if and only if $n_c > 0$, and since $F_{\mathbb{R}P^2}(2) = F_{\mathbb{R}P^2}(6) = 0$, we learn from equation (2.45) that $F_Y(2) = F_Y(6) = 0$ for any unorientable Y . So for those values of $N \bmod 8$, only orientable manifolds contribute after the sum over pin^- structures.

We will now begin to discuss how the connection between JT gravity and random matrix theory is modified by including unorientable surfaces and the pin^- sum. Suppose first that we want to compute the expectation value of $\langle Z_{\text{NS}}(\beta) \rangle$ or $\langle Z_{\text{R}}(\beta) \rangle$. We recall that these are computed in JT gravity by summing over contributions of two-manifolds Y whose boundary is a single circle X . In the present context, Y is a possibly unoriented two-manifold with a pin^- structure.

Then $\langle Z_{\text{R}}(\beta) \rangle$ remains zero, because a pin^- manifold with boundary always has an even number of boundary components of R type. If Y has just a single boundary component, it will be of NS type. On the other hand, $\langle Z_{\text{NS}}(\beta) \rangle$ is nonzero, and can receive contributions from unoriented as well as oriented manifolds. The dominant unoriented contribution comes from the manifold Y with the largest possible Euler characteristic, given that it is supposed to have one boundary component. This is a trumpet that ends on a crosscap (fig. 5); we will more briefly call it the crosscap spacetime.

Contribution of the crosscap spacetime. We will now evaluate the contribution of this manifold and compare to the random matrix ensembles identified above. In making this comparison, we will assume the purely bosonic JT gravity / RMT correspondence discussed in section 2.2, so that without any minus sign or sum over pin structures, the JT gravity partition function on the crosscap spacetime agrees with a GOE-like random matrix theory. There are two pin^- structures on the crosscap, with the same values $\eta = \pm 1/2$ as for \mathbb{RP}^2 . Hence the pin^- sum is just $F_{\mathbb{RP}^2}(N)$. The dominant crosscap contribution to $\langle Z_{\text{NS}}(\beta) \rangle$ is therefore as follows:

For $N = 0 \pmod 8$, since $F_{\mathbb{RP}^2}(0) = 2$, the crosscap contribution to $Z_{\text{NS}}(\beta)$ is two times the JT path integral for the crosscap spacetime, which we assume is equal to the GOE-like answer. In random matrix theory, for $N = 0 \pmod 8$, there are two independent GOE blocks, see table 1 below, so the factor of two matches.

For $N = 1 \pmod 8$, the pin^- sum is $F_{\mathbb{RP}^2}(1) = \sqrt{2}$. In random matrix theory, we have a single GOE block, see table 1 again, but the path integral and the trace in the Hilbert space are related by the factor of $\sqrt{2}$ in (2.8), so we find agreement.

For $N = 2 \pmod 8$, we have $F_{\mathbb{RP}^2}(2) = 0$, so the crosscap does not contribute. This agrees with table 1, since the $N = 2$ case consists of GUE blocks, for which there is no genus 1/2 contribution.

For $N = 3 \pmod 8$, the pin^- sum is $F_{\mathbb{RP}^2}(3) = -\sqrt{2}$. On the random matrix side, the only relevant difference from $N = 1 \pmod 8$ is that now $\hat{\mathbb{T}}^2 = -1$, so that we have to use GSE statistics instead of GOE. Again, this reverses the sign of the crosscap contribution.

For $N = 4 \pmod 8$, the pin^- sum is $F_{\mathbb{RP}^2}(4) = -2$. On the random matrix side, the only relevant difference from $N = 0 \pmod 8$ is that now $\mathbb{T}^2 = -1$, so that we have to use GSE statistics instead of GOE. This reverses the sign of the crosscap contribution.

The remaining cases follow by exchanging N with $8 - N \pmod 8$, so they do not add much.

Contribution of the double trumpet. Now we will discuss the leading contributions to the connected correlation functions $\langle Z_{\text{NS}}(\beta)^2 \rangle_c$ and $\langle Z_{\text{R}}(\beta)^2 \rangle$. These come from the same double trumpet studied in section 2.4.3, along with another topology that we will describe momentarily.

In section 2.4.3, we constructed the double trumpet Y by gluing together two trumpets Y_L and Y_R along their inner boundaries S_L and S_R (fig. 2). However, in a time-reversal invariant theory, we are free to make a reflection of S_R , reversing its orientation, before gluing it onto S_L . This makes a

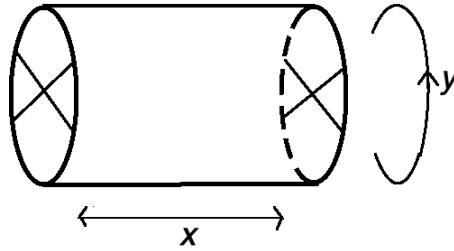


Figure 6: A Klein bottle can be constructed by closing off a cylinder at each end with a crosscap.

different manifold that we will call the twisted double trumpet \tilde{Y} . \tilde{Y} is not equivalent to Y by any diffeomorphism that acts trivially on its boundaries, so in the context of duality between an SYK-like model (or random matrix theory) and a bulk theory, we should consider \tilde{Y} and Y to be inequivalent.²²

In comparing $\exp(-i\pi N\eta/2)$ for different pin^- structures on Y or on \tilde{Y} , as usual, we can use locality to reduce to a simpler picture. A convenient way to use locality is to glue together the outer boundaries of Y or of \tilde{Y} . This produces from Y the two-torus of fig. 3, while from \tilde{Y} it produces a Klein bottle, which we will call KB.

A standard description of KB is that it can be obtained from the $x - y$ plane by dividing by two symmetries T_x and T_y , where

$$(2.47) \quad T_x(x, y) = (x + 1, y), \quad T_y(x, y) = (-x, y + 1).$$

It is the x direction here that parametrizes the circle S_L or S_R in the twisted double trumpet (or the external boundaries, to which S_L and S_R are homologous), since it is the x direction that is reflected by one of these operations (namely T_y). Another way to look at KB is also useful. The identifications $(x, y) \cong (x + 1, y)$ and $(x, y) \cong (-x, y + 1)$ let us restrict to $0 \leq x \leq 1/2$. At generic x , the equivalence relation on y is $y \cong y + 2$, but at the endpoints there is a further identification $y \cong y + 1$. This means that KB can be constructed from a cylinder $0 \leq x \leq 1/2$, $y \cong y + 2$ by closing it off with a crosscap at each end (fig. 6). This description implies that a complete pin^- sum of KB is just $F_{\text{KB}}(N) = F_{\mathbb{RP}^2}(N)^2$. But that is not quite what we need;

²²We could also rotate S_R (rather than reflecting it, or in addition) before gluing it to S_L . This is important in the path integral of JT gravity, but as it does not affect the topology, it is not visible in the topological field theory.

to understand contributions to $\langle Z_{\text{NS}}(\beta)^2 \rangle_c$ or $\langle Z_{\text{R}}(\beta)^2 \rangle$, we need to restrict the sum over pin^- structures on KB to those for which the pin^- structure on the boundaries of \tilde{Y} are of NS or R type, respectively. It turns out that these restricted pin^- sums are $F_{\text{KB}}^{\text{NS}}(N) = 2$, $F_{\text{KB}}^{\text{R}}(N) = 2 \cos(2\pi N/4)$ (this claim is explained at the end of the present section). Similarly, we can define the restricted pin^- sums for the torus. Since this manifold is orientable, pin^- structures are the same as spin structures, and we reduce to the case studied in section 2.4.3, which we can summarize as $F_T^{\text{NS}}(N) = 2$ and $F_T^{\text{R}}(N) = 1 + (-1)^N$.

These results imply that

$$(2.48) \quad F_T^{\text{NS}}(N) + F_{\text{KB}}^{\text{NS}}(N) = 4,$$

independent of N , while

$$(2.49) \quad F_T^{\text{R}}(N) + F_{\text{KB}}^{\text{R}}(N) = 1 + (-1)^N + 2 \cos 2\pi N/4 \\ = \begin{cases} 4 & \text{if } N = 0 \pmod 4 \\ 0 & \text{otherwise.} \end{cases}$$

To compare to random matrix theory, we will need one standard fact that we will borrow from section 4: the leading contribution to the connected correlator $\langle Z(\beta)^2 \rangle_c$ in the GOE or GSE-like ensembles is twice as big as in the GUE ensemble. More precisely, this is true if we define the GSE case to include both of each pair of degenerate eigenvalues, which is implicit in table 1.

If $N = 0, 4 \pmod 8$, then for either $\langle Z_{\text{NS}}(\beta)^2 \rangle_c$ or $\langle Z_{\text{R}}(\beta)^2 \rangle$, the pin^- sums of the double trumpet and the twisted double trumpet add to $2 + 2 = 4$, according to the above formulas. This multiplies what we would get by computing this correlator in bosonic JT gravity. This is consistent with random matrix theory, since each of the two GOE or GSE blocks in table 1 contribute twice the GUE answer, for a total factor of 4.

If N is odd, eqns. (2.48) and (2.49) imply that $\langle Z_{\text{NS}}(\beta)^2 \rangle_c$ is four times the GUE answer, and $\langle Z_{\text{R}}(\beta)^2 \rangle = 0$. Both are consistent with random matrix theory. From (2.8), the NS answer is $(\sqrt{2})^2 = 2$ times the GOE or GSE answer, so four times the GUE answer. Also, from (2.8), the R answer should be zero because for odd N the path integral in the Ramond sector vanishes.

If $N = 2, 6 \pmod 8$, the random matrix ensemble consists of two identical GUE-like blocks. The twofold degeneracy implies a factor of four in

$\langle Z_{\text{NS}}(\beta)^2 \rangle_c$, and since the two blocks have identical spectra, the Ramond-sector partition function is identically zero. Both facts are consistent with the pin^- sums (2.48) and (2.49).

Finally, to justify the claim about the restricted pin^- sums on KB, we first note that the four pin^- structures on KB can be presented by describing a fermion field λ on KB as a fermion field on \mathbb{R}^2 that satisfies

$$(2.50) \quad \begin{aligned} \lambda(x+1, y) &= (-1)^\alpha \lambda(x, y) \\ \lambda(-x, y+1) &= (-1)^\beta \gamma_x \lambda(x, y). \end{aligned}$$

Here $\alpha, \beta \in \{0, 1\}$ label the four choices of pin^- structure on KB, and γ_x is a gamma matrix (the pin^- condition means that $\gamma_x^2 = -1$, while pin^+ means $\gamma_x^2 = +1$). Since the x direction parametrizes the external boundaries of the twisted double trumpet (note the comment after eqn. (2.47)), this means that parallel transport of a fermion around the external boundary is controlled by $(-1)^\alpha$. Thus $\alpha = 1$ means that the external boundaries are of NS type, and $\alpha = 0$ means R type.

Now we view KB as a cylinder with a crosscap at $x = 0$ and another crosscap at $x = 1/2$. Eqn. (2.50) says that at $x = 0$, the pin^- structure is described by $\lambda(0, y+1) = (-1)^\beta \gamma_x \lambda(0, y)$. But at $x = 1/2$, the pin^- structure is instead described by $\lambda(1/2, y+1) = (-1)^{\alpha+\beta} \gamma_x \lambda(1/2, y)$. This says that if $\alpha = 0$, the pin^- structures on the two crosscaps are isomorphic, regardless of β , but if $\alpha = 1$, they are opposite.

This is all we need to compute the restricted pin^- sums. If the boundaries are of R type, we set $\alpha = 0$. The two crosscaps have isomorphic pin^- structures, each contributing $1/2$ or $-1/2$ to η , so the sum is $\eta = 1$ or $\eta = -1$. Therefore $F_{\text{KB}}^{\text{R}} = \exp(-i\pi N/2) + \exp(i\pi N/2) = 2 \cos(2\pi N/4)$, as claimed earlier. If the boundaries are of NS type, we set $\alpha = 1$. The two crosscaps have pin^- structures of opposite types. If one contributes $\pm 1/2$ to η , the other contributes $\mp 1/2$, so that overall $\eta = 0$. So $F_{\text{KB}}^{\text{NS}} = 2$, as also claimed earlier.

Reduction to the GOE, GUE, and GSE cases. At this point, we actually have enough information about pin^- structures to give a general reduction that shows consistency with random matrix theory, assuming it in the purely bosonic cases from section 2.2. Showing this requires several steps. First, one copies the steps in section 2.4.4 to reduce the $Z_{g,n,n_{\text{R}}}$ coefficient functions defined in (2.18) to combinations of $Z_{g,n}$ in the GOE, GUE, GSE-like ensembles. The predictions are given in the third column of table 1.

Before proceeding further, we note that in topological field theory, according to eqn. (2.49), amplitudes with boundaries of R type vanish unless

$N \bmod 8$	RMT class	Z_{g,n,n_R}	pin^- sum identity ($n_R = 0$)
0	$\begin{pmatrix} \text{GOE}_1 & 0 \\ 0 & \text{GOE}_2 \end{pmatrix}$	$2^{2g+n-2}(1 + (-1)^{n_R})Z_{g,n}^{\text{GOE}}$	$\sum_{\text{pin}^-} 1 = 2^{2g+n-1}$
1	GOE	$\begin{cases} 2^{g+n-1}Z_{g,n}^{\text{GOE}} & n_R = 0 \\ 0 & \text{else} \end{cases}$	$\sum_{\text{pin}^-} e^{-i\pi\eta/2} = 2^{g+n-1}$
2	$\begin{pmatrix} \text{GUE} & 0 \\ 0 & \text{GUE} \end{pmatrix}$	$\begin{cases} 2^{2g+2n-2}Z_{g,n}^{\text{GUE}} & n_R = 0 \\ 0 & \text{else} \end{cases}$	$\sum_{\text{pin}^-} e^{-i\pi\eta} = \begin{cases} 2^{2g+n-1} & n_c = 0 \\ 0 & \text{else} \end{cases}$
3	GSE	$\begin{cases} 2^{g+n-1}Z_{g,n}^{\text{GSE}} & n_R = 0 \\ 0 & \text{else} \end{cases}$	$\sum_{\text{pin}^-} e^{-3i\pi\eta/2} = 2^{g+n-1}(-1)^{n_c}$
4	$\begin{pmatrix} \text{GSE}_1 & 0 \\ 0 & \text{GSE}_2 \end{pmatrix}$	$2^{2g+n-2}(1 + (-1)^{n_R})Z_{g,n}^{\text{GSE}}$	$\sum_{\text{pin}^-} e^{-2i\pi\eta} = 2^{2g+n-1}(-1)^{n_c}$

Table 1: Random matrix classes as a function of $N \bmod 8$, for the cases with $T^2 = 1$ classically. Column two gives the RMT classes described in section 2.5.2 (for $N = 2$, the two GUE blocks are not independent but are exchanged by T). In the third column we follow the logic of section 2.4.4 to reduce the expansion coefficients Z_{g,n,n_R} to expansion coefficients in the ordinary GOE, GUE, GSE ensembles. In the final column, we give the pin^- sum identity (2.51) that implies this reduction in the bulk theory. The values of $N \bmod 8$ not listed here are obtained by $N \rightarrow 8 - N$.

$N = 0, 4 \bmod 8$. This agrees with the random matrix theory prediction in the third column of the table. If $N = 0, 4 \bmod 8$, the restricted pin^- sums for an NS or R boundary are the same, so from a topological field theory point of view, NS and R boundaries are equivalent except for the constraint that the total number of R boundaries is even. This is also a random matrix theory prediction, as stated in the third column in the table. Therefore, to complete the demonstration that topological field theory and random matrix theory make compatible predictions, it suffices to consider NS boundaries only.

Next, one uses gluing to derive the pin^- sum on a general surface Y . To do so, we start with a closed orientable surface Y_0 obtained from Y by removing the holes and crosscaps. Y_0 has genus $g_0 = g - \frac{1}{2}n_c$, where n_c is the number of crosscaps in Y . Because Y_0 is orientable, the pin^- sum is the same as a spin sum, and leads to 2^{2g_0} for even N and 2^{g_0} for odd N . Next, we glue in the n_c crosscaps, using (2.45). Finally, we cut n holes with boundaries S_j , and glue in n trumpets. As already explained, it suffices to assume the these are of NS type, in which case the sum over the “orthogonal” pin^- structure

gives a factor of 2 for each boundary. The pin^- sum is therefore

$$(2.51) \quad \frac{1}{2} 2^n 2^{2g_0} F_{\mathbb{RP}^2}(N)^{n_c}$$

for even N ; for odd N , it is obtained by replacing 2^{2g_0} with 2^{g_0} . Using $F_{\mathbb{RP}^2}(N) = 2 \cos 2\pi N/8$, this leads to the results stated in the last column of the table for the pin^- sums.

Comparing the last two columns in the table, we see that in all cases, the topological field theory prediction is in agreement with the reduction to GOE, GUE, GSE from the first step.²³

2.5.4. Cases with $\mathsf{T}^2 = (-1)^{\mathsf{F}}$ classically. Now we want to consider another symmetry group instead of $\mathbb{Z}_2^{\mathsf{T}} \times \mathbb{Z}_2^{\mathsf{F}}$. We still assume a time-reversal transformation T and a symmetry $(-1)^{\mathsf{F}}$ that distinguishes bosons and fermions. But now at the classical level we assume

$$(2.52) \quad \mathsf{T}^2 = (-1)^{\mathsf{F}}, \quad \left((-1)^{\mathsf{F}} \right)^2 = 1.$$

So $\mathsf{T}^4 = 1$ and T generates a group that we call $\mathbb{Z}_4^{\mathsf{T}}$. How to construct an SYK-like model with such a symmetry group was explained in section 2.5.1. In particular the number N of elementary fermions must be even.

This case is actually simpler than the previous case of $\mathbb{Z}_2^{\mathsf{T}} \times \mathbb{Z}_2^{\mathsf{F}}$, so we can be more brief. For even N , we described in section 2.5.2 an irreducible representation of the Clifford algebra with $N/2$ real generators and $N/2$ imaginary ones. Call the even generators χ_k , $k = 1, \dots, N/2$, and the odd ones $\tilde{\chi}_k$, $k = 1, \dots, N/2$. Define the time-reversal transformation

$$(2.53) \quad \mathsf{T} = \frac{\chi_1 + \tilde{\chi}_1}{\sqrt{2}} \frac{\chi_2 + \tilde{\chi}_2}{\sqrt{2}} \dots \frac{\chi_{N/2} + \tilde{\chi}_{N/2}}{\sqrt{2}} \mathsf{K}.$$

This has been chosen so that

$$(2.54) \quad \mathsf{T} \chi_k \mathsf{T}^{-1} = \tilde{\chi}_k, \quad \mathsf{T} \tilde{\chi}_k \mathsf{T}^{-1} = -\chi_k.$$

A short computation gives

$$(2.55) \quad \mathsf{T}^2 = i^{N^2/4} (-1)^{\mathsf{F}},$$

²³For $N = 2 \bmod 8$, one has to remember that the JT gravity path integral in a time-reversal-invariant theory is 2^{n-1} times the path integral in the GUE-like theory without time-reversal symmetry; see the end of section 2.2.

with

$$(2.56) \quad (-1)^F = i^{N/2} \chi_1 \tilde{\chi}_1 \chi_2 \tilde{\chi}_2 \cdots \chi_{N/2} \tilde{\chi}_{N/2}.$$

Up to an overall sign, the phase of $(-1)^F$ has been fixed to ensure $((-1)^F)^2 = 1$.

If $N = 0 \pmod 4$, we have the expected relations (2.52). The random matrix symmetry class of a Hamiltonian with these symmetries consists of two blocks for the \pm eigenvalues of $(-1)^F$. Acting on the $+$ subspace, $T^2 = 1$, so we have GOE-like statistics. Acting on the $-$ subspace, $T^2 = -1$, so we have GSE-like statistics.

If $N = 2 \pmod 4$, we have instead

$$(2.57) \quad T^2 = i(-1)^F.$$

If we absorb this factor of i in the definition of $(-1)^F$, we would get $((-1)^F)^2 = -1$. So there is no way to avoid an anomaly, though it can be moved around. Since T commutes with T^2 and anticommutes with i , eqn. (2.57) implies the important fact that if $N = 2 \pmod 4$, T anticommutes with $(-1)^F$:

$$(2.58) \quad T(-1)^F = -(-1)^F T.$$

This implies that the two blocks of the Hamiltonian are exchanged by T , but are not constrained individually. So we have two GUE-like blocks with identical eigenvalues.

Thus we have found an anomaly that depends on the value of $N \pmod 4$. There is no more subtle anomaly, since by perturbing the Hamiltonian in a suitable fashion, one can remove two pairs of fermions, say $\chi_1, \tilde{\chi}_1$ and $\chi_2, \tilde{\chi}_2$, from the low energy spectrum. One simply adds the “mass” term $\Delta H = -im(\chi_1 \chi_2 + \tilde{\chi}_1 \tilde{\chi}_2)$. For large m , this set of four fermions is removed from the low energy effective theory, reducing N by 4, without breaking any symmetry.

Since N is even and the anomaly only depends on $N \pmod 4$, this is a \mathbb{Z}_2 -valued anomaly. The bulk dual of a theory with \mathbb{Z}_4^T symmetry is a theory in which (in Euclidean signature) one sums over pin^+ structures on two-manifolds. Topological field theories that can be built by summing over pin^+ structures in suitable ways have a \mathbb{Z}_2 classification, which matches what we need.

The necessary invariant is as follows. Let Y be a two-manifold, not necessarily oriented, with a pin^+ structure. On an unorientable two-manifold

Y , it is not possible to consider a chiral fermion field, but we can of course consider a non-chiral₂ two-component fermion field λ . Now we can define a \mathbb{Z}_2 -valued invariant $\tilde{\zeta}$ by imitating the definition of the invariant ζ in section 2.4.3. We simply define $\tilde{\zeta}$ to be the mod 2 index of the nonchiral Dirac operator on Y , that is the dimension mod 2 of the space of solutions of the Dirac equation $\not{D}\lambda = 0$. This is a topological invariant, as discussed in section 3 of [76]. On a two-manifold without boundary, one can make two topological field theories by summing over pin^+ structures with or without a factor of $(-1)^{\tilde{\zeta}}$. As in other cases we have discussed, $\tilde{\zeta}$ has the necessary locality properties to justify this statement. On an orientable manifold, the two theories are equivalent because $\tilde{\zeta} = 0$ if Y is orientable.²⁴ This is as one would expect, because \mathbb{Z}_4^T symmetry forces N to be even.

On a two-manifold with boundary, there is a problem with the definition of $\tilde{\zeta}$: although one can place a local boundary condition on the Dirac equation and then count its zero-modes mod 2, such a local boundary condition is not invariant under time-reversal symmetry, that is under a reflection of the boundary. So $(-1)^{\tilde{\zeta}}$ has an anomaly on a manifold with boundary. One can show that this anomaly matches the anomaly that we found above for N congruent to 2 mod 4. One way to show this is to imitate arguments given in [72] for the analogous statement about ζ , which was invoked in section 2.4.3.

Accordingly, we propose that an SYK-like model with the sort of time-reversal symmetry that we have discussed here should be compared to JT gravity with a sum over pin^+ structures, with or without a factor of $(-1)^{\tilde{\zeta}}$, depending on whether N is congruent to 2 or 0 mod 4. Equivalently, we can in general include a factor $(-1)^{\frac{N}{2}\tilde{\zeta}}$.

Since $\tilde{\zeta} = 0$ if Y is orientable, it follows that the difference between the two theories only comes into play for unorientable Y . This is as we would expect, since N is even in the present context, and anomalies that can be seen on orientable manifolds only depend on N mod 2.

For illustration and because it will be useful, we will compute $\tilde{\zeta}$ for a pin^+ structure on a Klein bottle KB. Such structures can be described exactly as we described pin^- structures in eqn. (2.50),

$$(2.59) \quad \begin{aligned} \lambda(x+1, y) &= (-1)^\alpha \lambda(x, y) \\ \lambda(-x, y+1) &= (-1)^\beta \gamma_x \lambda(x, y), \end{aligned}$$

²⁴If Y is orientable, the dimension of the space of solutions of the Dirac equation is $n_+ + n_-$, where n_+ and n_- are the dimensions of the spaces of solutions with positive or negative chirality. But $n_+ = n_-$, because complex conjugation reverses the fermion chirality in two dimensions. So $n_+ + n_-$ is always even.

with the only difference being that now $\gamma_x^2 = +1$ rather than -1 . A zero-mode of the Dirac equation on KB is simply a constant mode of λ that obeys the conditions (2.59). If $\alpha = 1$, there is no such mode. But if $\alpha = 0$, then since γ_x is a 2×2 matrix with eigenvalues ± 1 , there is, regardless of β , precisely one zero-mode. So $\tilde{\zeta} = 0$ if $\alpha = 1$ and $\tilde{\zeta} = 1$ if $\alpha = 0$.

Now let us view KB as a cylinder with a crosscap at each end, as in fig. 6. There are two pin^+ structures on each crosscap. By the same argument as given for pin^- structures in section 2.5.3, $\alpha = 0$ means that the pin^+ structures on the two crosscaps are isomorphic and $\alpha = 1$ means that they are opposite. So changing α from 0 to 1 flips the pin^+ structure on one of the two crosscaps, regardless of β . Since changing α also changes $\tilde{\zeta}$, we learn that flipping the pin^+ structure on one crosscap changes the value of $\tilde{\zeta}$.

By locality, this statement is universal and not limited to the particular case of a Klein bottle. This leads to a simple statement about the sum over pin^+ structures on any manifold. Let $G_Y(N)$ be the sum of pin^+ structures on a manifold Y , weighted by $(-1)^{\frac{N}{2}\tilde{\zeta}}$. If Y is unorientable, it can be obtained by gluing a crosscap onto some other manifold Y' . Pairs of pin^+ structures on Y that differ by flipping the pin^+ structure on this crosscap will, by locality, make canceling contributions to $G_Y(N)$ if $N = 2 \pmod{4}$. So for $N = 2 \pmod{4}$, only orientable manifolds contribute to the partition function after the sum over pin^+ structures. This is consistent with the random matrix class identified above, consisting of two identical GUE-like blocks.

On the other hand, for $N = 0 \pmod{4}$, unorientable manifolds do contribute, and lead to something new. A difference between pin^+ structures and pin^- structures is that if Y is a pin^+ manifold with boundary, it can have an odd number of boundary components of R type. In fact, the basic crosscap spacetime of fig. 4 – a trumpet that ends on a crosscap – is an example. The crosscap spacetime has two pin^+ structures, both of which are of R type on the boundary. Hence this spacetime contributes to $\langle Z_R(\beta) \rangle$, not to $\langle Z_{NS}(\beta) \rangle$. (This is opposite from the pin^- case; a pin^- structure on the crosscap spacetime is of NS type on the boundary.) This is consistent with random matrix theory: for $N = 0 \pmod{4}$, the Hamiltonian consists of a GOE block with $(-1)^F = 1$ and a GSE block with $(-1)^F = -1$. The crosscap (or genus $1/2$) contributions in these two random matrix ensembles are equal in magnitude but opposite in sign, so they cancel in $Z_{NS}(\beta)$ and add in $Z_R(\beta)$.

One can go on and look at the connected correlations functions $\langle Z_{NS}(\beta)^2 \rangle_c$ and $\langle Z_R(\beta)^2 \rangle_c$. The lowest order contributions will come from the double trumpet and the twisted double trumpet.

$N \bmod 4$	RMT class	Z_{g,n,n_R}	pin ⁺ structure sum identity
0	$\begin{pmatrix} \text{GOE} & 0 \\ 0 & \text{GSE} \end{pmatrix}$	$2^{2g+n-2} (Z_{g,n}^{\text{GOE}} + (-1)^{n_R} Z_{g,n}^{\text{GSE}})$	$\sum_{\text{pin}^+} 1 = 2^{2g+n-2} (1 + (-1)^{n_R+n_c})$
2	$\begin{pmatrix} \text{GUE} & 0 \\ 0 & \text{GUE} \end{pmatrix}$	$\begin{cases} 2^{2g+2n-2} Z_{g,n}^{\text{GUE}} & n_R = 0 \\ 0 & \text{else} \end{cases}$	$\sum_{\text{pin}^+} (-1)^{\tilde{\zeta}} = \begin{cases} 2^{2g+n-1} & n_c = 0 \\ 0 & n_R = 0 \\ 0 & n_c \neq 0 \end{cases}$

Table 2: Reductions to GOE, GUE, GSE as in table 1, but now for the cases with $T^2 = (-1)^F$ classically. For $N = 2 \bmod 4$, T exchanges the two GUE blocks.

For $N = 0 \bmod 4$, the double trumpet and twisted double trumpet have two pin⁺ structures, each of which contributes +1, regardless of whether the boundaries are of NS or R type. Thus the contributions to $\langle Z_{\text{NS}}(\beta)^2 \rangle_c$ and $\langle Z_{\text{R}}(\beta)^2 \rangle_c$ are equal. The cross-correlator $\langle Z_{\text{R}}(\beta) Z_{\text{NS}}(\beta) \rangle_c$ is nonzero, but will receive contributions only from spacetimes with odd Euler characteristic. The leading term would be from \mathbb{RP}^2 with two holes removed.

For $N = 2 \bmod 4$, the double trumpet and twisted double trumpet contributions to the correlator $\langle Z_{\text{R}}(\beta) \rangle_c$ cancel. The cancellation happens because, with the boundaries being of R type, one has $(-1)^{\tilde{\zeta}} = 1$ for a pin⁺ structure on the double trumpet and $(-1)^{\zeta} = -1$ for a pin⁺ structure on the twisted double trumpet. To prove this, using locality, one can glue the two boundaries together, so that the double trumpet and the twisted double trumpet become a torus and a Klein bottle. The torus is orientable, so it has $\zeta = 0$ for any pin⁺ structure, and on the Klein bottle, since we are now considering pin⁺ structures with $\alpha = 0$ in the sense of eqn (2.59), we have $\tilde{\zeta} = 1$. This cancellation is expected in random matrix theory because $Z_{\text{R}}(\beta)$ vanishes identically, without fluctuations.

As in the previous cases, one can make a general argument that pin⁺ structure sums ensure that random matrix theory answer agrees with the bulk. See table 2 for the necessary pin⁺ identities, which follow from the analog of (2.51). Concretely, the above considerations show that for $N = 0 \bmod 4$, the orthogonal pin⁺ sum for any NS or R boundary is 2, so we get eqn. (2.51) with $F_{\mathbb{RP}^2}(N)$ replaced by $G_{\mathbb{RP}^2}(0) = 2$. In pin⁺, crosscaps are in the R sector, so instead of imposing that the number of R boundaries should be even, we impose that the number of R boundaries plus the number of crosscaps should be even. This accounts for the result claimed in the table for $N = 0 \bmod 4$. For $N = 2 \bmod 4$, the orthogonal pin⁺ sum vanishes for an R boundary, and also $G_{\mathbb{RP}^2}(2) = 0$. So we restrict to $n_R = n_c = 0$. The orthogonal pin⁺ sum still gives a factor of 2 for every NS boundary. In

contrast to the pin^- case, the pin^+ sum on an orientable surface Y_0 of genus g_0 is 2^{2g_0} for any N . These facts lead to the result shown in the table for $N = 2 \pmod 4$.

2.6. Including $\mathcal{N} = 1$ supersymmetry

2.6.1. The supersymmetric SYK model. The supersymmetric SYK model, with minimal or $\mathcal{N} = 1$ supersymmetry, is constructed [81] by starting with a quantum mechanical system of N Majorana fermions and introducing a random supercharge

$$(2.60) \quad Q = i^{\frac{\hat{q}-1}{2}} \sum_{a_1 a_2 \dots a_{\hat{q}}} \tilde{j}_{a_1 a_2 \dots a_{\hat{q}}} \psi_{a_1} \psi_{a_2} \dots \psi_{a_{\hat{q}}}.$$

Here \hat{q} is an odd integer ≥ 3 . The Hamiltonian is defined as

$$(2.61) \quad H = Q^2.$$

The assertion that the model is supersymmetric just means that $[H, Q] = 0$.

If Q is a homogeneous function of the elementary fermions, as assumed in eqn. (2.60), then the supersymmetric SYK model inevitably has $\mathbb{Z}_2^{\text{T}} \times \mathbb{Z}_2^{\text{F}}$ symmetry (possibly with a quantum anomaly). Indeed, if T is a time-reversal symmetry that commutes with the elementary fermions, then

$$(2.62) \quad \text{T}Q = (-1)^{(\hat{q}-1)/2} Q \text{T}.$$

Thus Q is either even or odd under T , and therefore H is even. However, by allowing Q to be a sum of terms with different values of \hat{q} , one can break the time-reversal symmetry.

Regardless of how Q is constructed, $\text{T}^2 = (-1)^{\text{F}}$ is not compatible with $\mathcal{N} = 1$ supersymmetry. If $\text{T}^2 = (-1)^{\text{F}}$, then fermionic operators transform under T in even-dimensional representations, but in the case of $\mathcal{N} = 1$ supersymmetry, Q is a fermionic operator that is in a 1-dimensional representation.

As in a nonsupersymmetric SYK model, we can exchange T with $\text{T}(-1)^{\text{F}}$, and this will change the sign of the anomaly. But it also reverses the sign in eqn. (2.62). This sign is important in the random matrix classification, and also in the bulk supergravity. So unlike the nonsupersymmetric case, it is not true that models differing by $N \rightarrow 8 - N$ will be equivalent. Rather, a model defined by a pair (N, \hat{q}) will be equivalent to a model with parameters

$(8 - N, \widehat{q} + 2)$. So, for example, it suffices to consider only the case $\widehat{q} = 1 \bmod 4$ if we consider all values of $N \bmod 8$, as in Table 4 of the next section.

2.6.2. Random matrix classes including supersymmetry. The random matrix classification of the supersymmetric SYK model was studied in [66–68]. We will give a self-contained description. First consider the case without time-reversal symmetry, so the only global symmetry is the supercharge Q and a possible $(-1)^F$ symmetry. If N is even, there is a $(-1)^F$ symmetry. The two blocks of states with $(-1)^F = 1$ or with $(-1)^F = -1$ each have dimension $L = 2^{N/2-1}$. The unitary group acting on either block is $U(L)$, so the full group of unitary transformations that commutes with $(-1)^F$ is $G = U(L) \times U(L)$. Q anticommutes with $(-1)^F$, so in a basis in which $(-1)^F$ is block diagonal, Q has only off-diagonal blocks:

$$(2.63) \quad Q = \begin{pmatrix} 0 & C \\ C^\dagger & 0 \end{pmatrix}.$$

Here C is a complex $L \times L$ matrix obeying no constraint, and C^\dagger is its hermitian adjoint. C transforms as a bifundamental of $U(L) \times U(L)$. Thus the random matrix statistics will correspond to one of the Altland-Zirnbauer ensembles that were described in section 2.1.1. If N is odd, then there is no $(-1)^F$ symmetry. The symmetry group G reduces to $U(L)$ (with now $L = 2^{(N-1)/2}$) and Q is simply a hermitian matrix, governed by GUE statistics.

Returning to the case that there is a $(-1)^F$ symmetry, there is a possible generalization that does not occur in the SYK model but that is natural from a random matrix point of view. We will interpret it in JT supergravity in section 5.5. One can assume that the blocks of bosonic and fermionic states have different dimensions $L + \nu$ and L , for some integer ν . The symmetry group is then $U(L + \nu) \times U(L)$; C is still a bifundamental. This corresponds to a supersymmetric matrix ensemble in which supersymmetry is unbroken, the value of the supersymmetric index being $\text{Tr}(-1)^F = \nu$. The same generalization is possible in the other examples with bifundamentals.

Now we consider the case with time-reversal symmetry, but with N even so that Q has the block structure in eqn. (2.63). Then the role of T is to reduce the symmetry group and place a constraint on C . If $N = 0 \bmod 8$, then $T^2 = 1$ and the symmetry group that commutes with T and with $(-1)^F$ is $G = O(L) \times O(L)$. C is a bifundamental of this group. If $N = 4 \bmod 8$, then $T^2 = -1$, the symmetry group is $G = \text{Sp}(L) \times \text{Sp}(L)$, and C is again a bifundamental of G . These are two more of the Altland-Zirnbauer ensembles. If $N = 2$ or $6 \bmod 8$, then T exchanges the two blocks, and the analysis is slightly more subtle. Since T exchanges the two blocks, the symmetry group

$N \bmod 2$	symmetry	$\lambda \rightarrow 0$ enhancement	$\lambda_1 \rightarrow \lambda_2$ enhancement	α	β	γ	bulk TFT
0	$U(L) \times U(L)$ bif.	$U(1) \rightarrow U(1) \times U(1)$	$U(1) \times U(1) \rightarrow U(2)$	1	2	2	$(-1)^\zeta$
1	$U(L)$ adjoint	$U(1) \rightarrow U(1)$	$U(1) \times U(1) \rightarrow U(2)$		2	$\sqrt{2}$	1

Table 3: Random matrix symmetry classification for the supercharge Q in the $\mathcal{N} = 1$ super SYK model with N fermions and with no time-reversal symmetry. The matrix field is a bifundamental of $U(L) \times U(L)$ or an adjoint of $U(L)$, as indicated in the second column. In the third and fourth columns, we give the symmetry enhancements described in section 2.1.2. The dimensions of these enhancements directly give α and β , also shown. The variable γ is the degeneracy of eigenvalues, multiplied by $\sqrt{2}$ for the case of odd N , see (5.43). In the first row, we have an Altland-Zirnbauer (α, β) ensemble, and in the second row, we have an ordinary $\beta = 2$ (GUE-like) ensemble (so no value of α is given).

that commutes with \mathbb{T} (and with $(-1)^F$) is just $G = U(L)$: one can make an arbitrary unitary transformation of the upper block, accompanied by a \mathbb{T} -conjugate transformation of the lower block. Let \mathcal{H}_+ and \mathcal{H}_- be the subspaces of the super SYK Hilbert space \mathcal{H} corresponding to the eigenvalue $+1$ or -1 of $(-1)^F$, respectively. Since \mathcal{H}_+ and \mathcal{H}_- are exchanged by the antiunitary symmetry \mathbb{T} , they are naturally dual. Hence instead of thinking of C as a linear map from \mathcal{H}_+ to \mathcal{H}_- , we can think of it as a bilinear map $\mathcal{H}_+ \otimes \mathcal{H}_+ \rightarrow \mathbb{C}$ or in other words as a second rank tensor. But we need to decide if C is a symmetric or antisymmetric second rank tensor, and this point is somewhat subtle. Denoting as $\text{Spin}(N)$ the group that rotates the N elementary fermions of the SYK model, the elementary fermions transform in the vector representation V of this group, and the Hilbert spaces \mathcal{H}_+ and \mathcal{H}_- transform, respectively, as the positive and negative chirality spinor representations, which we will call S_+ and S_- . Suppose that Q is homogeneous of degree \hat{q} in the elementary fermions. Then Q and C transform in the representation $\wedge^{\hat{q}}V$ (the antisymmetric product of \hat{q} copies of V). To decide if C is a symmetric or antisymmetric tensor, we just need to know if $\wedge^{\hat{q}}V$ appears symmetrically or antisymmetrically in $S_+ \otimes S_+$. For $\hat{q} = 1 \bmod 4$, the group theory answer is that $\wedge^{\hat{q}}V$ appears symmetrically if $N = 2 \bmod 8$ and antisymmetrically if $N = 6 \bmod 8$; these statements are reversed if $\hat{q} = 3 \bmod 4$. So for $N = 2$ or $6 \bmod 8$, we get two more Altland-Zirnbauer ensembles; which one occurs for $N = 2 \bmod 8$ and which for $N = 6 \bmod 8$ depends on \hat{q} .

N	symmetry	$\lambda \rightarrow 0$ enhancement	$\lambda_i \rightarrow \lambda_j$ enhancement	T_+^2	T_-^2	α	β	γ	bulk
0	$O(L) \times O(L)$ bif.	$1 \rightarrow 1$	$1 \rightarrow O(2)$	+1	+1	0	1	2	$e^{i\pi\eta/2}$
1	$O(L)$ symm.	$1 \rightarrow 1$	$1 \rightarrow O(2)$	+1			1	$\sqrt{2}$	1
2	$U(L)$ symm.	$1 \rightarrow U(1)$	$1 \rightarrow U(1)$	+1	-1	1	1	2	$e^{-i\pi\eta/2}$
3	$Sp(L)$ symm.	$U(1) \rightarrow Sp(2)$	$U(1) \times U(1) \rightarrow U(2)$		-1	2	2	$2\sqrt{2}$	$e^{-2i\pi\eta/2}$
4	$Sp(L) \times Sp(L)$ bif.	$Sp(2) \rightarrow Sp(2) \times Sp(2)$	$Sp(2) \times Sp(2) \rightarrow Sp(4)$	-1	-1	3	4	4	$e^{-3i\pi\eta/2}$
5	$Sp(L)$ anti.	$Sp(2) \rightarrow Sp(2)$	$Sp(2) \times Sp(2) \rightarrow Sp(4)$	-1			4	$2\sqrt{2}$	$e^{-4i\pi\eta/2}$
6	$U(L)$ anti.	$SU(2) \rightarrow U(2)$	$SU(2) \times SU(2) \rightarrow Sp(4)$	-1	+1	1	4	4	$e^{-5i\pi\eta/2}$
7	$O(L)$ anti.	$O(2) \rightarrow O(2)$	$O(2) \times O(2) \rightarrow U(2)$		+1	0	2	$2\sqrt{2}$	$e^{-6i\pi\eta/2}$

Table 4: Random matrix symmetry classification for the supercharge Q in the $\mathcal{N} = 1$ super SYK model with N fermions and with $\widehat{q} = 1 \pmod{4}$. The first column is $N \pmod{8}$. The second column indicates the symmetry group and whether the matrix is a bifundamental of a product group or a symmetric or antisymmetric second rank tensor. The next two columns indicate the symmetry enhancement when an eigenvalue goes to zero or two eigenvalues become equal (here 1 denotes the trivial group). The operators T_{\pm} are antiunitary operators built out of T and $(-1)^F$ so that they commute (+) or anticommute (-) with the supercharge Q ; a blank means that there is no such operator. For the cases where α is blank, the matrix ensemble for Q is a standard Dyson β ensemble. The others are Altland-Zirnbauer (α, β) ensembles. The values of α, β follow from the symmetry enhancements, as explained in section 2.1.2. γ is the degeneracy, multiplied by $\sqrt{2}$ when N is odd. The last column shows a factor that must be included in the sum over pin^- structures. (The corresponding table for $\widehat{q} = 3 \pmod{4}$ is obtained from this one by $N \rightarrow 8 - N \pmod{8}$.)

With time-reversal and odd N , there is only one block in Q . The symmetry group is reduced from $U(L)$ to $O(L)$ if $T^2 = 1$, which happens for $N = 1$ or $7 \pmod{8}$, or to $Sp(L)$ if $T^2 = -1$, as happens for $N = 3$ or $5 \pmod{8}$ (see section 2.5.2 for these statements). As detailed momentarily, T either commutes or anticommutes with Q , depending on N and \widehat{q} . If Q commutes with T , then Q is governed by GOE statistics for $T^2 = 1$ or GSE statistics for $T^2 = -1$. But if T anticommutes with Q , then Q is governed by an ensemble that we have not yet encountered. If $T^2 = 1$ and T anticommutes with Q , then $Q = iM$, where M is a real antisymmetric second rank tensor of $O(L)$ (equivalently, an element of the adjoint representation of $O(L)$). If $T^2 = -1$,

then Q transforms as a symmetric tensor²⁵ of $\mathrm{Sp}(L)$ (equivalently, the adjoint representation of $\mathrm{Sp}(L)$). These are the last two Altland-Zirnbauer ensembles.

We give a summary of this classification in tables 3 and 4. In these tables we also list the values of α and β , along with a variable γ that characterizes the degeneracy of levels in Q^2 and that will be used in section 5. In table 4 for the time-reversal-invariant case, we also give the values of \mathbb{T}_+^2 and \mathbb{T}_-^2 . These are defined as the antiunitary operators that either commute (+) or anticommute (-) with Q . They are related to \mathbb{T} and $\mathbb{T}(-1)^F$ in a way that depends on N .

2.6.3. Bulk description including supersymmetry. A bulk description of the supersymmetric SYK model involves JT supergravity at low energies. This motivates a conjecture that pure JT supergravity, together with the appropriate topological field theory, should be dual to matrix integrals of the types just discussed. To determine the topological field theory, we match the anomalies in the SYK model.

In the case without time-reversal symmetry, the potential anomaly is in $(-1)^F$. In the bulk we sum over spin structures on orientable manifolds, with the weights 1 or $(-1)^\zeta$. Naively, one would identify the trivial theory with the non-anomalous random matrix class (even N). However, there is an important subtlety. Ordinary JT gravity includes a ‘‘Schwarzian’’ mode, which is a bosonic mode that propagates along the boundary. In JT supergravity, the analog is a Schwarzian supermultiplet, a supermultiplet that propagates along the boundary [30, 81]. In particular, the Schwarzian supermultiplet includes a Majorana fermion. This boundary mode contributes to the anomaly.

To see how this affects things, suppose we start with a boundary theory without the $(-1)^F$ anomaly. We look for a bulk dual description in terms of JT supergravity, including the Schwarzian supermultiplet, plus a bulk topological field theory. The Schwarzian supermultiplet by itself has the $(-1)^F$

²⁵Let $M_{ij} = M_{ji}$ be a symmetric tensor of $\mathrm{Sp}(L)$. Since the fundamental representation of $\mathrm{Sp}(L)$ is not real, the components of M_{ij} are not real, but M can satisfy an $\mathrm{Sp}(L)$ -invariant reality condition $\overline{M}^{ij} = -\varepsilon^{ii'} \varepsilon^{jj'} M_{i'j'}$, where ε^{ij} is the invariant antisymmetric tensor of $\mathrm{Sp}(L)$ (which we take to be real). Equivalently, with ε and M viewed as matrices, the reality condition on M is $M^\dagger = \varepsilon M \varepsilon$. Then the matrix $Q^i_j = \varepsilon^{ik} M_{kj}$ is hermitian, since $Q^\dagger = (\varepsilon M)^\dagger = M^\dagger \varepsilon^\dagger = \varepsilon M \varepsilon \varepsilon^\dagger = \varepsilon M = Q$, where we use the fact that $\varepsilon \varepsilon^\dagger = 1$. The hermitian matrix Q is an element of the adjoint representation of $\mathrm{Sp}(L)$. It is odd under \mathbb{T} , since $\mathbb{T}(Q) = \varepsilon \overline{Q} \varepsilon^{-1} = \varepsilon \overline{\varepsilon M \varepsilon}^{-1} = \overline{M} \varepsilon = -\varepsilon M = -Q$. We used $\overline{\varepsilon} = \varepsilon$, $\varepsilon^2 = -1$.

anomaly, so the bulk topological field theory must cancel it. We conclude that the bulk description of the non-anomalous (even N) theory should include $(-1)^\zeta$, and the bulk description of the anomalous (odd N) theory should be the trivial theory.

The logic for the case with time-reversal symmetry is similar. In the bulk, we sum over orientable and unorientable manifolds. To match the $\mathbb{Z}_2^T \times \mathbb{Z}_2^F$ symmetry, we sum over pin^- structures, and the possible topological field theory is $\exp(-i\pi\eta N'/2)$ with N' an integer mod 8. To relate this N' to the N of SYK, we have to keep track of the contribution to the anomaly of the Schwarzian supermultiplet. The fermion in this multiplet transforms under T like the supercharge Q . That means that its contribution to the anomaly is just the sign in eqn. (2.62), or $(-1)^{(\hat{q}-1)/2}$.

Now, suppose we start with a system of N Majorana fermions. We want to match the anomaly in the bulk. The Schwarzian supermultiplet contributes $(-1)^{(\hat{q}-1)/2}$ to the anomaly, so the bulk topological field theory must contribute what remains, or

$$(2.64) \quad N' = N - (-1)^{(\hat{q}-1)/2}.$$

Note that unlike the cases studied in sections 2.4 and 2.5, here we cannot use spin structure sums to reduce the different cases to the GOE, GUE, GSE cases. The bulk reason is simply that JT supergravity contains fermions, so the path integral depends on the spin or pin structure in a nontrivial way.

As a final comment, one might wonder if JT supergravity plus topological field theory would be unable to distinguish N' from $8 - N' \bmod 8$, on the grounds that exchanging N' with $8 - N'$ could be compensated by exchanging a pin^- bundle \mathcal{P} with the complementary one \mathcal{P}' . This is not the case, because the path integral of JT supergravity is sensitive to the difference between \mathcal{P} and \mathcal{P}' , basically because it knows how the gravitino field transforms under T .

3. Torsion

Our next goal is to understand the path integral of JT gravity or supergravity on an unorientable two-manifold. For this we will have to understand the “torsion.” Since JT gravity or supergravity can be formulated as a BF theory, we start by explaining how torsion is related to BF theory.

Torsion in topology originally meant the combinatorial torsion of Reidemeister [82]. Roughly, on a triangulated manifold on which a flat background gauge field is given, one defines the torsion (or R-torsion or combinatorial

torsion) as a certain product of determinants which is somewhat miraculously independent of how the manifold was triangulated. One can calculate the torsion explicitly by triangulating a manifold in a very crude way with only a few lattice points. On the other hand, by taking a very fine triangulation, one can approximate a continuum limit. That last fact, in hindsight, could be a clue that torsion has a quantum field theory interpretation. Our interest in the torsion is related to the fact that it has a continuum limit, but in our calculations later, we will take advantage of the fact that it is possible to use a crude triangulation.

Long after Reidemeister's construction, a differential geometric or analytic version of the torsion was defined by Ray and Singer [83]. Analytic torsion is defined in terms of a product of determinants of differential operators. The differential operators that appear are the Laplace-like operators that arise in quantizing gauge theories. With hindsight, this could serve as a clue that this form of the torsion is related to a one-loop approximation to gauge theory. Indeed, it was shown by A. Schwarz that Ray-Singer analytic torsion can arise as the partition function of what is now called BF theory [51]. Later, this result was an ingredient in understanding the one-loop approximation to Chern-Simons theory in three dimensions [84].

Ray and Singer introduced ζ -function regularization in order to define their determinants. Combinatorial torsion can be viewed as a lattice regularization of the same determinants. From a physical point of view, one might expect two regularizations of the same theory to give equivalent results, possibly after adjusting some local counterterms. In the present case, no such counterterms arise in odd dimensions, but there are some in even dimensions.²⁶ In hindsight, this might have provided a reason to suspect that combinatorial and analytic torsion are equivalent. The equivalence, which was conjectured by Ray and Singer, was proved by Cheeger [85] and Müller [86], originally for compact gauge groups. The proof was later generalized to a larger class of groups that includes noncompact but semi-simple Lie groups such as $SL(2, \mathbb{R})$ [87–89].

²⁶In the two-dimensional case relevant to the present paper, there is one possible counterterm that is consistent with the fact that the analytic and combinatorial torsions are both topological invariants. This is a factor $\exp(w\chi)$, where w is a constant and χ is the Euler characteristic of a manifold. Such a factor, in general, should be expected whenever one compares two formulations of BF theory. The analytic and combinatorial torsions, however, have been defined in a way that avoids such a factor.

When there are no zero-modes, the torsion is a number. More generally, under favorable conditions,²⁷ such as prevail in the two-dimensional case that will be of interest here, the torsion is instead a measure on the moduli space of flat connections. On an even-dimensional orientable manifold (without boundary), the torsion is often said to be “trivial.” Triviality means that the torsion is 1 if there are no zero-modes. When the torsion is a measure on moduli space, the “triviality” means that this measure can be defined by a simple classical formula, without any quantum correction. However, on an unorientable two-manifold (or an orientable one with boundary), there is no such “triviality” and there is a quantum correction. For compact gauge groups, volumes of moduli spaces of flat connections on orientable and unorientable two-manifolds were computed in [90] using the combinatorial definition of the torsion.²⁸ Here we will perform an analogous computation for gauge group $SL(2, \mathbb{R})$ and its supersymmetric analog $OSp(1|2)$. There is a technical difference in the way we will describe the answer. In [90], inspired by what was known about two-dimensional Yang-Mills theory [91], the torsion was written in terms of a sum over characters of the gauge group (which provide a basis for the space of physical states). While such a representation might be possible for $SL(2, \mathbb{R})$ or $OSp(1|2)$, it turns out to be very convenient instead to calculate directly with Fenchel-Nielsen or length-twist coordinates.

3.1. Analytic torsion

Let us first review the relationship between BF theory and the torsion. We consider a theory with a gauge group G whose Lie algebra \mathfrak{g} admits an invariant, nondegenerate quadratic form that we will denote as Tr . The fields will be a gauge field A with field strength $F = dA + A \wedge A$, and an adjoint-valued spin zero field B . The action on a two-manifold Y is

$$(3.1) \quad I = -i \int_Y \text{Tr} BF.$$

This is well-defined even for unorientable Y , provided that we take B to be a pseudoscalar. The path integral of this theory is one-loop exact, since the

²⁷Quantizing a BF theory in higher dimensions requires introducing a hierarchy of “ghosts for ghosts,” as explained in [51]. When the higher order ghosts have zero-modes, the torsion does not have a simple interpretation as a measure on a moduli space of classical solutions.

²⁸Some important early ideas on this and related matters had been developed in unpublished work by S. Axelrod.

integral over B gives a delta function setting $F = 0$:

$$(3.2) \quad \int \mathcal{D}B \exp(-I) = \int \mathcal{D}B \exp(i \int \text{Tr} BF) = \delta(F).$$

After gauge-fixing, the delta function localizes A on the moduli space of flat connections modulo gauge transformations. The resulting measure on the moduli space is given by a product of determinants – a ghost determinant that arises in gauge fixing, and a second determinant that arises in integrating over A after gauge-fixing with the help of the $\delta(F)$ in eqn. (3.2).

Though it is possible to continue in precisely this way, we will take a small detour so as to be able to present the derivation in a more familiar language. Let us add to the action I a B^2 term with a small coefficient ε :

$$(3.3) \quad I_\varepsilon = -i \int_Y \text{Tr} BF - \frac{\varepsilon}{2} \int_Y d^2x \sqrt{g} \text{Tr} B^2,$$

where g is a Riemannian metric on Y . Integrating out B , we get

$$(3.4) \quad I'_\varepsilon = -\frac{1}{2\varepsilon} \int d^2x \sqrt{g} \text{Tr} F^2.$$

This is simply weakly coupled Yang-Mills theory. The one-loop approximation (which is not exact in Yang-Mills theory, but is exact in the limit $\varepsilon \rightarrow 0$, in which we return to the original BF theory) is given by a standard formula. Suppose that we are expanding around a background flat connection A_0 , say with $A = A_0 + a$ where a describes the fluctuations. Let d be the exterior derivative and $D = d + [A_0, \cdot]$ its gauge-covariant extension, mapping adjoint-valued q -forms to adjoint-valued $(q + 1)$ -forms for $q = 0, 1$. Its adjoint D^* maps q -forms to $(q - 1)$ -forms for $q = 2, 1$. The Laplacian acting on adjoint-valued differential forms is $\Delta = D^*D + DD^*$. We write Δ_q for the Laplacian acting on adjoint-valued q -forms. Then a standard computation shows that after gauge-fixing, the one-loop path integral of the Yang-Mills theory gives a ratio of determinants

$$(3.5) \quad Z_{1,A_0} = \frac{\det' \Delta_0}{\sqrt{\det' \Delta_1}}.$$

Here in a standard approach, the denominator comes from the Gaussian integral over a after gauge-fixing, and the numerator is the ghost determinant. The symbol \det' represents a determinant in a subspace orthogonal to the zero-modes. If Δ_1 has zero-modes, this means that the classical solution A_0 around which we are expanding is not unique but represents a

point in a moduli space \mathcal{M} of classical solutions; in this case, Z_{1,A_0} must be interpreted as a measure on \mathcal{M} . If Δ_0 has zero modes, this means that the classical solution A_0 leaves unbroken a positive-dimensional subgroup of G and the Fadde'ev-Popov gauge-fixing should be discussed more carefully. This typically does not occur in two dimensions for nonabelian G .

To understand the formula (3.5), we need to take into account the following identity:

$$(3.6) \quad \det' \Delta_1 = \det' \Delta_0 \det' \Delta_2.$$

The origin of this identity is as follows. The Laplacian Δ commutes with D and D^* , so if $\psi \in \Omega^0$ is an eigenstate of Δ_0 , then $D\psi \in \Omega^1$ is an eigenstate of Δ_1 with the same eigenvalue. Likewise if $\chi \in \Omega^2$ is an eigenstate of Δ_2 , then $D^*\chi \in \Omega^1$ is an eigenstate of Δ_1 with the same eigenvalue. Hodge theory says every nonzero eigenstate of Δ_1 can be identified in a unique way with an eigenstate of Δ_0 or of Δ_2 with the same eigenvalue, leading to the identity (3.6). In fact, Hodge theory gives an orthogonal decomposition of Ω^1 as

$$(3.7) \quad \Omega^1 = \Omega_0^1 \oplus \Omega_2^1 \oplus \Omega_{\text{harm}}^1,$$

where Ω_0^1 consists of states of the form $D\psi$, $\psi \in \Omega^0$, Ω_2^1 consists of the states $D^*\chi$, $\chi \in \Omega^2$, and Ω_{harm}^1 is the space of zero-modes, which are known as harmonic forms.

Using the identity (3.6), we get

$$(3.8) \quad Z_{1,A_0} = \frac{\sqrt{\det' \Delta_0}}{\sqrt{\det' \Delta_2}}.$$

Now we can see why the one-loop correction is trivial if Y is orientable (and without boundary²⁹), but not otherwise. On an orientable manifold of dimension D , one has the Hodge star operator mapping n -forms to $(D-n)$ -forms and ensuring that $\det' \Delta_n = \det' \Delta_{D-n}$. So in particular on an orientable two-manifold Y without boundary, $\det' \Delta_2 = \det' \Delta_0$, and the 1-loop correction is trivial. On an unorientable manifold, $\det' \Delta_2 \neq \det' \Delta_0$, and the 1-loop correction is nontrivial.

Now let us discuss what we do with Z_{1,A_0} when it is nontrivial. To define the path integral of gauge theory, one starts formally with a Riemannian

²⁹If Y has a boundary, one has to pick a boundary condition, which will typically not be invariant under the Hodge star operator.

metric on the fluctuation field \mathbf{a} ,

$$(3.9) \quad \|\mathbf{a}\|^2 = - \int_Y d^2x \sqrt{g} \text{Tr} \mathbf{a}_i \mathbf{a}_j g^{ij},$$

which induces a Riemannian measure. This measure (and a similar one for the ghosts and antighosts) is used, formally, in defining the functional integral. The Riemannian metric on the space of all \mathbf{a} 's induces, in particular, a Riemannian metric on the space of zero-modes. This in turn determines a Riemannian metric on the moduli space \mathcal{M} of classical solutions, and therefore this space gets a Riemannian measure μ_0 . The one-loop determinant corrects this measure to $\mu = \frac{\sqrt{\det' \Delta_0}}{\sqrt{\det' \Delta_2}} \mu_0$, and the partition function of BF theory for a compact non-abelian gauge group is obtained by integrating this measure over the moduli space \mathcal{M} of flat connections:

$$(3.10) \quad Z_{BF} = \int_{\mathcal{M}} d\mu = \int_{\mathcal{M}} d\mu_0 \frac{\sqrt{\det' \Delta_0}}{\sqrt{\det' \Delta_2}}.$$

(For a group such as $SL(2, \mathbb{R})$, one also wants to divide by the mapping class group of Y in order to get a finite answer.) If Y is orientable, the measure that is used here is the elementary one μ_0 whose definition did not require BF theory, but if Y is unorientable, the quantum correction $\sqrt{\det' \Delta_0 / \det' \Delta_2}$ is important.³⁰

3.2. Combinatorial torsion

It is possible to imitate some of these formulas for a triangulated two-manifold Y and more generally for one that has been built by gluing together polygons (fig. 7). This will lead us to the definition of the Reidemeister torsion. By a q -cell we will mean one of the q -dimensional building blocks in a

³⁰A further subtlety arises if the gauge group has a nontrivial center \mathcal{Z} . If we are on a manifold without boundary, because \mathcal{Z} acts trivially on any flat connection, the Fadde'ev-Popov recipe tells us to supplement the ghost determinant with a factor of $1/\#\mathcal{Z}$, where $\#\mathcal{Z}$ is the number of elements of \mathcal{Z} . For $SL(2, \mathbb{R})$, this would be a factor of $1/2$. In our applications, we are usually interested in a two-manifold Y with a nontrivial boundary. In that context, we only allow gauge transformations that are trivial along the boundary, so a constant gauge transformation by an element of \mathcal{Z} is not allowed and we do not divide by $\#\mathcal{Z}$. On the contrary, if the boundary of Y has h components, then we have to sum over 2^{h-1} inequivalent classes of gauge bundles on Y that would become equivalent if we allowed gauge transformations that are nontrivial on the boundary. This factor was explained in section 2.5.4 in the context of the double trumpet, which is an example with $h = 2$.

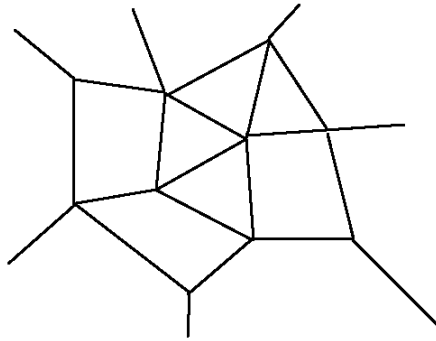


Figure 7: A two-manifold built by gluing together polygons (not necessarily triangles) on their boundaries.

covering by polygons (a vertex if $q = 0$, an edge if $q = 1$, and a polygon if $q = 2$). We write \mathcal{S}_q for the set of all these q -cells. Such a covering is sometimes called a cell decomposition. A somewhat more thorough explanation of the following can be found, for example, in [90], section 4.³¹

In the combinatorial approach to torsion, it is convenient to use a language that is dual to some of what we have said so far. Instead of q -forms, one considers functions on q -cells (every q -form defines such a function, namely its integral over a given q -cell) and instead of the exterior derivative D mapping q -forms to $(q + 1)$ -forms, one has the boundary operator ∂ mapping q -cells to $(q - 1)$ -cells. The q -cells are considered to be oriented, and if W is a q -cell, then its boundary ∂W is an oriented sum of $(q - 1)$ -cells.

Suppose that on a two-manifold Y , we are given a flat G -bundle. Let E be the associated bundle in the adjoint representation. Any particular oriented q -cell w is contractible, so when restricted to w , the bundle E has a trivialization by a vector space \mathcal{V}_w consisting of covariantly constant sections³² of E over w . Then we define for $q = 0, 1, 2$ a vector space that is

³¹There was a much more thorough explanation in a 1988 preprint by D. Johnson, “A Geometric Form of Casson’s Invariant and Its Connection With Reidemeister Torsion.” Unfortunately this article was unpublished and appears to be unavailable online. Any reader with a copy is invited to share it.

³²It can happen in a given covering by polygons that two or more vertices (if $q = 1, 2$) or edges (if $q = 2$) of a given q -cell are glued together. We ignore any such boundary identifications in defining \mathcal{V}_w . One may say that \mathcal{V}_w consists of covariantly constant sections of E over the interior of w , which is always contractible, regardless of what gluing occurs on the boundary of w . If a given edge or vertex appears more than once in ∂w , then we add the different contributions in defining ∂s for $s \in \mathcal{V}_w$.

the direct sum of the \mathcal{V}_w for all $w \in \mathcal{S}_q$:

$$(3.11) \quad \mathcal{C}_q = \oplus_{w \in \mathcal{S}_q} \mathcal{V}_w.$$

Now we define maps $\partial : \mathcal{C}_q \rightarrow \mathcal{C}_{q-1}$ as follows. Let w be a q -cell and suppose that geometrically the boundary of w is $\partial w = \cup_i y_i$, with y_i being $(q - 1)$ -cells. Then one defines $\partial : \mathcal{V}_w \rightarrow \oplus_i \mathcal{V}_{y_i}$ by $\partial s = \oplus_i s|_{y_i}$. More generally, remembering that all the cells are oriented and their orientations may not be compatible, we write the boundary of w as $\partial w = \cup_i (-1)^{\lambda_i} y_i$, where $\lambda_i = 0, 1$ and a minus sign means that the orientation of y_i is not the one that is induced by the orientation of w . Then we define $\partial s = \oplus_i (-1)^{\lambda_i} s|_{y_i}$. Having defined maps $\partial : \mathcal{V}_w \rightarrow \mathcal{C}_{q-1}$ for all $w \in \mathcal{S}_q$, we just sum over w to get $\partial : \mathcal{C}_q \rightarrow \mathcal{C}_{q-1}$.

The Lie algebra \mathfrak{g} of G can be given a G -invariant and translation-invariant measure, which is unique up to a constant multiple. For example, such a measure could be deduced from a G -invariant inner product on \mathfrak{g} , which determines a flat G -invariant metric on \mathfrak{g} and hence a G -invariant Riemannian measure. As each \mathcal{V}_w is a copy of \mathfrak{g} , each \mathcal{V}_w gets a measure, and therefore each $\mathcal{C}_q = \oplus_{w \in \mathcal{S}_q} \mathcal{V}_w$ gets a measure, which we will call α_q . These measures will be important in what follows. Though there was an arbitrary choice of the underlying measure on \mathfrak{g} , it can be shown that a change in this measure will just multiply the combinatorial torsion by $\exp(w\chi)$, where w is a constant and χ is the Euler characteristic of Y . The proof of this assertion uses the fact that $\chi = n_0 - n_1 + n_2$, where n_q is the number of q -cells in the cell decomposition of Y , together with a simple scaling argument using the definition (3.15) of the torsion.

Up to now, we have defined a linear transformation $\partial : \mathcal{C}_q \rightarrow \mathcal{C}_{q-1}$ for $q = 1, 2$; when we want to be more precise, we distinguish $\partial_2 : \mathcal{C}_2 \rightarrow \mathcal{C}_1$ and $\partial_1 : \mathcal{C}_1 \rightarrow \mathcal{C}_0$. Because the boundary of a boundary vanishes, these maps satisfy $\partial_1 \partial_2 = 0$. That lets one define homology groups $H_q(Y, E)$, which will play a role shortly. The homology groups are defined as follows: $H_2(Y, E)$ is the kernel of $\partial_2 : \mathcal{C}_2 \rightarrow \mathcal{C}_1$; $H_1(Y, E)$ is $\ker \partial_1 / \text{im } \partial_2$, that is it is the quotient of the kernel of ∂_1 by the image of ∂_2 ; and $H_0(Y, E)$ is $\mathcal{C}_0 / \partial_1 \mathcal{C}_1$, that is it is the quotient of \mathcal{C}_0 by the image of ∂_1 . A standard fact in topology is that the vector spaces $H_q(Y, E)$ do not depend on the specific cell decomposition of Y that was used to define them.

At this point, a quick way to define the Reidemeister or combinatorial torsion is to set $\widehat{\Delta}_2 = \partial_2^* \partial_2$, $\widehat{\Delta}_0 = \partial_1 \partial_1^*$, and then we can define the Reidemeister torsion:

$$(3.12) \quad \tau^{\mathbf{R}} = \frac{\sqrt{\det' \widehat{\Delta}_2}}{\sqrt{\det' \widehat{\Delta}_0}}.$$

This is exactly in parallel with eqn. (3.8) except that the roles of q and $2 - q$ have been exchanged. This has happened because we have employed a dual description based on the operator ∂ which reduces q by 1 rather than the operator d which increases q by 1, used in the analytic approach. The remarkable property of $\tau^{\mathbf{R}}$ is that it does not depend on the combinatorial description, so we can use a crude triangulation, leading to simple formulas, or a fine one, approaching a continuum limit. The main step in showing that the choice of a cell decomposition does not matter is to show that subdividing one of the polygons does not change $\tau^{\mathbf{R}}$. See, for example, section 4 of [90].

Both conceptually and for computational purposes, a slightly different formula for $\tau^{\mathbf{R}}$ is convenient. In describing this, we will first assume that the homology groups $H_q(Y, E)$ vanish.

We will use the measures α_q on the vector spaces \mathcal{C}_q . In general, if V is a vector space and α is a measure on V , then to any basis v_1, v_2, \dots, v_k of V , α assigns a number $\alpha(v_1, v_2, \dots, v_k)$, which is a linear function of each of the v_i . One can think of $\alpha(v_1, v_2, \dots, v_k)$ as the volume of a parallelepiped that has vertices at the origin in V as well as the points $v_1, v_2, \dots, v_k \in V$. If $M : V \rightarrow V$ is a linear transformation, then

$$(3.13) \quad \alpha(Mv_1, Mv_2, \dots, Mv_k) = |\det M| \alpha(v_1, v_2, \dots, v_k).$$

This formula involves $|\det M|$, not $\det M$, because α is a measure, not a differential form of top degree.

Now if $H_2(Y, E) = 0$, and s_1, s_2, \dots, s_{n_2} is any basis \mathcal{C}_2 , then the vectors $\partial s_1, \partial s_2, \dots, \partial s_{n_2}$ are linearly independent in \mathcal{C}_1 , so they can be completed to a basis $\partial s_1, \partial s_2, \dots, \partial s_{n_2}, t_1, t_2, \dots, t_{n_1-n_2}$ of \mathcal{C}_1 . The conditions $H_1(Y, E) = H_0(Y, E) = 0$ mean that $\partial t_1, \partial t_2, \dots, \partial t_{n_1-n_2}$ provide a basis of \mathcal{C}_0 . Therefore, we have bases of \mathcal{C}_2 , \mathcal{C}_1 , and \mathcal{C}_0 , and using the measures α_q on these vector spaces, we can define

$$(3.14) \quad \tau^{\mathbf{R}} = \frac{\alpha_2(s_1, s_2, \dots, s_{n_2}) \alpha_0(\partial t_1, \partial t_2, \dots, \partial t_{n_1-n_2})}{\alpha_1(\partial s_1, \partial s_2, \dots, \partial s_{n_2}, t_1, t_2, \dots, t_{n_1-n_2})}.$$

This formula does not depend on the choices of the s_i or the t_j , basically because each s_i and each t_j appears in both the numerator and the denominator in eqn. (3.14), so a rescaling of the s_i or the t_j , or a more general linear

transformation of the s 's or of the t 's, does not affect τ^R . As we have already remarked, τ^R also does not depend on the specific cell decomposition that was used.

However, for our application, we need to relax the assumption that the homology is trivial. For some purposes, we can maintain the assumption $H_0(Y, E) = H_2(Y, E) = 0$, but we should not assume that $H_1(Y, E) = 0$. On the contrary, $H_1(Y, E)$ is the cotangent bundle to the moduli space \mathcal{M} of flat bundles on Y , at the point corresponding to a given flat bundle E . Suppose that $\dim H_1(Y, E) = \dim \mathcal{M} = r$. Then we should modify the above description as follows. To the vectors $\partial s_1, \partial s_2, \dots, \partial s_{n_2}$, all of which are annihilated by ∂_1 since $\partial_1 \partial_2 = 0$, we should adjoin r more basis vectors u_1, u_2, \dots, u_r that are annihilated by ∂_1 and provide a basis of $H_1(Y, E)$. Then $\partial s_1, \partial s_2, \dots, \partial s_{n_2}, u_1, u_2, \dots, u_r$ are a basis of the kernel of ∂_1 . We complete this to a basis of \mathcal{C}_1 by adding vectors $t_1, \dots, t_{n_1-n_2-r}$ which are not annihilated by ∂_1 , after which as before (since we assume $H_0(Y, E) = 0$), $\partial t_1, \partial t_2, \dots, \partial t_{n_1-n_2-r}$ provide a basis of \mathcal{C}_0 . So we simply modify the definition of τ^R by including the u_i among the basis vectors of \mathcal{C}_1 :

$$(3.15) \quad \tau^R = \frac{\alpha_2(s_1, s_2, \dots, s_{n_2}) \alpha_0(\partial t_1, \partial t_2, \dots, \partial t_{n_1-n_2-r})}{\alpha_1(\partial s_1, \partial s_2, \dots, \partial s_{n_2}, u_1, u_2, \dots, u_r, t_1, t_2, \dots, t_{n_1-n_2-r})}.$$

As before, τ^R is independent of the s_i and the t_j , but it definitely does depend on the u_k , since they only appear in the denominator. In fact, if we rescale one of the u 's by a factor λ , τ^R is multiplied by a factor of λ^{-1} . More generally, since the u_k appear in the denominator, τ^R transforms in a change of basis as the *inverse* of a measure on $H_1(Y, E)$. The inverse of a measure on a vector space can be regarded as a measure on its dual space. In the present context, $H_1(Y, E)$ is the cotangent bundle of the moduli space \mathcal{M} of flat connections on Y , and its dual space³³ is $H^1(Y, E)$, the tangent bundle to \mathcal{M} . A measure on the tangent space to \mathcal{M} at every point in \mathcal{M} determines a measure on \mathcal{M} . In other words, τ^R is a measure on \mathcal{M} .

In a combinatorial approach, the partition function of BF theory is the volume of \mathcal{M} computed with this measure:

$$(3.16) \quad Z_{BF} = \int_{\mathcal{M}} d\tau^R.$$

³³The distinction between the tangent and cotangent bundle of \mathcal{M} is only important if Y is unorientable, because if Y is orientable, \mathcal{M} has a symplectic structure that does not depend on any choice of a metric on Y , and this can be used to identify the tangent and cotangent bundles of \mathcal{M} .

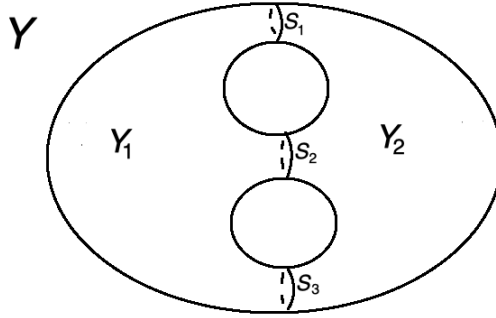


Figure 8: A genus 2 surface Y built by gluing two three-holed spheres Y_1 and Y_2 along boundary circles S_1, S_2, S_3 .

If $H^0(Y, E) \neq 0$, then we need to add more basis vectors v_1, \dots, v_k of \mathcal{C}_0 , and the definition becomes

$$(3.17) \quad \tau^R = \frac{\alpha_2(s_1, s_2, \dots, s_{n_2})\alpha_0(\partial t_1, \partial t_2, \dots, \partial t_{n_1-n_2-r}, v_1, v_2, \dots, v_k)}{\alpha_1(\partial s_1, \partial s_2, \dots, \partial s_{n_2}, u_1, u_2, \dots, u_r, t_1, t_2, \dots, t_{n_1-n_2-r})}.$$

From the foregoing, it is clear that τ^R can be effectively calculated, using a simple covering of Y by polygons. But to get a useful result, it is important to know the gluing law for the torsion.

3.3. Gluing

In general, a complicated two-manifold can be built by gluing together simple building blocks. For an example, an oriented surface Y of genus ≥ 2 can be built by gluing together three-holed spheres as in fig. 8.

From the point of view of BF theory, if Y is a Riemann surface with boundary ∂Y , one can perform the path integral over the fields on Y while keeping fixed the boundary values of the gauge field A on ∂Y . This path integral will give a function $\Psi(A)$ of the boundary values. This function defines an element in a Hilbert space \mathcal{H} of quantum states that live on ∂Y . If ∂Y is a union of several circles $S_i, i = 1, \dots, n$, then $\mathcal{H} = \otimes_{i=1}^n \mathcal{H}_i$, where \mathcal{H}_i is a Hilbert space associated to S_i . Suppose Y_1 and Y_2 are two Riemann surfaces with boundary and we want to glue them along some or all of their boundary components to make a surface Y which itself may or may not have a boundary. We can compute the BF path integral on Y by first computing BF path integrals on Y_1 and Y_2 separately, to generate Hilbert space states,

after which, for any circle on which we want to glue Y_1 and Y_2 , we take an inner product of the corresponding factor in the Hilbert spaces. Roughly, we “trace out” the Hilbert space factor associated to the circle on which gluing is supposed to occur.

In the example of fig. 8, we could calculate the BF path integral on Y all at once. Alternatively, we could calculate the path integral on Y_1 and Y_2 , keeping fixed the gauge field A on the three circles S_1, S_2, S_3 that are the common boundary of Y_L and of Y_R . If we write A_i for the restriction of A to S_i , then the BF path integral on Y_1 or on Y_2 with fixed boundary values generates a state $\Psi_1(A_1, A_2, A_3)$ or $\Psi_2(A_1, A_2, A_3)$. If we multiply the two states together and integrate over A_1, A_2, A_3 , we get the inner product (Ψ_1, Ψ_2) , which is simply the BF path integral on Y . The fact that the computation can be split up into pieces in this fashion is an important aspect of the locality of quantum field theory.

A similar cut and paste procedure is possible in the combinatorial approach to the torsion. However, the most convenient approach involves one difference in detail. First of all, suppose that Y is a two-manifold with boundary, and that we describe it combinatorially by gluing together polygons. The procedure described in section 3.2 is still valid even if Y has a boundary; τ^R as defined in eqn. (3.14) is a measure on the moduli space of flat connections on Y , even if Y has a boundary. In particular, let S be a circle that is part of the boundary of Y , and let a be a gauge-invariant function of the holonomy of A around S . For example, for $G = \text{SL}(2, \mathbb{R})$, we can take a to be (in a sense we make precise later) the logarithm of the holonomy. Then a is a function on the moduli space and we can loosely write a measure on \mathcal{M} , suppressing other variables, as $f(a)da$.

Now suppose Y_1 and Y_2 are two-manifolds with a common boundary circle S on which we want to glue them together. Now we have two moduli spaces \mathcal{M}_1 and \mathcal{M}_2 of flat connections on Y_1 and Y_2 respectively. Working on Y_1 , the logarithm of the holonomy of a flat connection around S is then a function on \mathcal{M}_1 that we will call a_1 . Similarly, working on Y_2 , the corresponding logarithm is a function on \mathcal{M}_2 that we will call a_2 . The torsion on Y_1 or on Y_2 gives a measure that we can schematically call $\tau_{Y_1} = f(a_1)da_1$ or $\tau_{Y_2} = g(a_2)da_2$, ignoring other variables. (Y_1 and Y_2 might be different topologically, so the functions f and g and the other variables that are being suppressed can be quite different in the two cases.)

When we glue Y_1 and Y_2 along S to make a two-manifold Y , what sort of gluing law for the torsion might we one hope for, analogous to what happens in BF theory? Naively, one might hope that after setting $a_1 = a_2$, we would get $\tau_Y^R = \tau_{Y_1}^R \tau_{Y_2}^R$. But this cannot be right, because after setting $a_1 = a_2$ and

writing, say, a for the common value, the product $\tau_{Y_1}^R \tau_{Y_2}^R = f(a)g(a)(da)^2$ is not a measure because it has an extra factor of da . To get a measure which might equal τ_Y^R , we need a way to remove one factor of da .

Intuitively, what is wrong is that since S is part of Y_1 and also part of Y_2 , when we simply take a product $\tau_{Y_1}^R \tau_{Y_2}^R$, we are double counting the degrees of freedom that live in S . To get the correct result, we have to remove one copy of S , which is accomplished by dividing by τ_S^R , the torsion of a flat connection on S . τ_S^R is defined exactly as in section 3.2, except that as S is a 1-manifold, it can be built from 1-cells and 0-cells, with no 2-cells. (We will go into more detail when we actually calculate τ_S^R .) The correct gluing law for the Reidemeister or combinatorial torsion, whenever Y_1 and Y_2 are glued along S to make Y , is³⁴

$$(3.18) \quad \tau_Y^R = \tau_{Y_1}^R \frac{1}{\tau_S^R} \tau_{Y_2}^R.$$

We will see the power of this formula when we actually perform computations.

Dividing by τ_S^R actually solves two problems. First, as already explained, it removes a surplus factor of da . But it also provides an extra factor that would be missing in the simple product $\tau_{Y_1}^R \tau_{Y_2}^R$. A flat bundle on Y can be restricted to Y_1 and to Y_2 , but it is not uniquely determined by those restrictions. To build \mathcal{M}_Y , the moduli space of flat connections on Y , out of the corresponding \mathcal{M}_{Y_1} and \mathcal{M}_{Y_2} , we have to set $a_1 = a_2$, as already discussed. But that is not all. On Y , there can be a new modulus, a relative “twist parameter” between fields on Y_1 and on Y_2 . For $G = \mathrm{SL}(2, \mathbb{R})$, the twist parameter in question is the “twist” variable in Fenchel-Nielsen or length-twist coordinates; see section 3.4.1. If we denote such a relative twist parameter as ϱ , then to build a measure on \mathcal{M}_Y from a product of measures on \mathcal{M}_{Y_1} and on \mathcal{M}_{Y_2} , we need to add a factor $d\varrho$ as well as removing a factor da . As we will see, dividing by τ_S^R accomplishes both of these tasks.

3.4. Computations for $G = \mathrm{SL}(2, \mathbb{R})$

3.4.1. Flat connections and hyperbolic geometry. JT gravity in the absence of time-reversal symmetry is related to BF theory with gauge group $\mathrm{PSL}(2, \mathbb{R})$ in the case of a purely bosonic theory, or $\mathrm{SL}(2, \mathbb{R})$ in the case of

³⁴A mathematical reference on the gluing law for the analytic torsion in its most general form is [92]. An explanation of the gluing law for the combinatorial torsion can be found, for example, in section 4 of [90].

a theory with fermions and spin structures.³⁵ The difference is that a flat $SL(2, \mathbb{R})$ connection encodes the spin structure as well as the geometry, while a flat $PSL(2, \mathbb{R})$ connection only encodes the geometry.

We may or may not want to consider a theory with fermions, but in any event, the torsion, with which we are concerned here, is not sensitive to the difference between $PSL(2, \mathbb{R})$ and $SL(2, \mathbb{R})$, basically because it is a one-loop effect of fields with values in the adjoint representation. We will use the language of $SL(2, \mathbb{R})$ because it is more convenient to write 2×2 matrices, not just because we might want to study a theory with fermions.

Actually, the relation of JT gravity to BF theory of $SL(2, \mathbb{R})$ refers to just one component³⁶ of the moduli space of flat $SL(2, \mathbb{R})$ connections. This is the component we are interested in, though some of our considerations can be applied to other components.

Three-holed spheres are a basic example, because (with a few exceptions that have non-negative Euler characteristic) any oriented two-manifold Y can be built by gluing together three-holed spheres. So let us discuss the hyperbolic geometry of a three-holed sphere.

If Y is a three-holed sphere, then every flat $SL(2, \mathbb{R})$ connection on Y (in the appropriate component, but we will not keep stating this qualification in what follows) is associated to a hyperbolic metric on Y , with constant scalar curvature $R = -2$, and with geodesic boundaries. The lengths of the three geodesic boundaries are arbitrary positive real numbers, so the relevant moduli space for a three-holed sphere (ignoring the spin structure, about which more in a moment) is a simple product \mathbb{R}_+^3 of three copies of \mathbb{R}_+ (the set of positive real numbers). In particular, the moduli space has dimension 3.

What is the connection between the length or circumference of a geodesic boundary component and the corresponding flat $SL(2, \mathbb{R})$ connection? This question has a rather simple answer. The monodromy of the flat connection around a boundary component of length ℓ is conjugate to

$$(3.19) \quad U_\ell^\pm = \pm \begin{pmatrix} e^{\ell/2} & 0 \\ 0 & e^{-\ell/2} \end{pmatrix}.$$

³⁵ $SL(2, \mathbb{R})$ is the group of 2×2 real matrices of determinant 1, and $PSL(2, \mathbb{R})$ is its quotient by $\{\pm 1\}$.

³⁶The group $SL(2, \mathbb{R})$ is contractible to $U(1)$, so an $SL(2, \mathbb{R})$ bundle over a topological space has a first Chern class, just like a $U(1)$ bundle. A *flat* $SL(2, \mathbb{R})$ bundle over a Riemann surface Y of genus $g > 1$ can have any value of the first Chern class from $g - 1$ to $-(g - 1)$. The component that is related to hyperbolic geometry is the component of first Chern class $g - 1$.

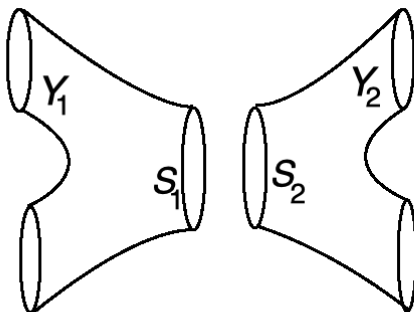


Figure 9: Y_1 and Y_2 are three-holed spheres with geodesic boundaries S_1 and S_2 . The only necessary condition for being able to glue S_1 onto S_2 , producing a smooth hyperbolic surface Y , is that S_1 and S_2 should have the same circumference. If the gluing is possible, then one can make an arbitrary rotation of S_1 relative to S_2 , by a “twist” parameter ϱ , prior to the gluing.

The sign, which we do not see if we project to $\mathrm{PSL}(2, \mathbb{R})$, determines the spin structure; a spin structure of NS or R type around the circle in question corresponds to a $-$ or $+$ sign, respectively. We see that

$$(3.20) \quad \mathrm{Tr} U_\ell^\pm = \pm 2 \cosh \ell/2.$$

Conversely, any element of $\mathrm{SL}(2, \mathbb{R})$ whose trace is $\pm 2 \cosh \ell/2$ is conjugate to U_ℓ^\pm . So the conjugacy class that contains U_ℓ^+ or U_ℓ^- is (for all $\ell \neq 0$) of codimension 1 in $\mathrm{SL}(2, \mathbb{R})$ and hence of dimension 2.

Given this, we can do a simple dimension counting and check the claim that the moduli space of flat $\mathrm{SL}(2, \mathbb{R})$ connections on the three-holed sphere has dimension 3. Let U, V, W be the monodromies around the three holes; they are subject to one condition $UVW = 1$. The space of all triples $U, V, W \in \mathrm{SL}(2, \mathbb{R})$ is nine-dimensional; after imposing the condition $UVW = 1$ and dividing by an overall conjugation by an arbitrary element of $\mathrm{SL}(2, \mathbb{R})$, we are left with a moduli space of dimension $9 - 3 - 3 = 3$. One can take the three moduli to be the three boundary lengths or equivalently the three traces $\mathrm{Tr} U, \mathrm{Tr} V, \mathrm{Tr} W$.

It is not difficult to describe explicitly the monodromies of a family of flat connections with variable boundary lengths on a three-holed sphere Y .

We can take, for example,

$$(3.21) \quad U = \delta_a \begin{pmatrix} e^{a/2} & \kappa \\ 0 & e^{-a/2} \end{pmatrix}, \quad V = \delta_b \begin{pmatrix} e^{-b/2} & 0 \\ 1 & e^{b/2} \end{pmatrix}.$$

depending on three real parameters a, b, κ . Here a and b are length parameters; the geodesics in question have lengths $|a|$ and $|b|$. And δ_a and δ_b are 1 or -1 for a spin structure of Ramond or NS type. To get $UVW = 1$, the third monodromy must be

$$(3.22) \quad W = V^{-1}U^{-1} = \delta_a \delta_b \begin{pmatrix} e^{-(a-b)/2} & -\kappa e^{b/2} \\ -e^{-a/2} & e^{(a-b)/2} + \kappa \end{pmatrix}.$$

This gives a flat $SL(2, \mathbb{R})$ connection for any real κ , but it is only associated with a hyperbolic metric on Y with geodesic boundaries if $2 \cosh \frac{1}{2}(a - b) + \kappa < -2$. In that case, the third length is $|c|$ where $\text{Tr } W = 2\delta_c \cosh c/2$ (where δ_c controls the spin structure of the third geodesic), and

$$(3.23) \quad 2 \cosh \frac{c}{2} = -\kappa - 2 \cosh \frac{a - b}{2}.$$

We have used the fact that $\delta_a \delta_b \delta_c = -1$ for any compatible set of spin structures (since the number of Ramond boundaries is always even). To get a hyperbolic three-holed sphere (rather than a flat $SL(2, \mathbb{C})$ connection in a different component of the moduli space) we must chose κ consistent with $\cosh \frac{c}{2} > 1$. Some more detail on this flat connection is described in appendix D.3.

Now let Y_1 and Y_2 be hyperbolic two-manifolds with geodesic boundary components S_1 and S_2 , respectively. Suppose that we want to make a hyperbolic manifold Y by gluing $S_1 \subset Y_1$ to $S_2 \subset Y_2$. An obvious necessary condition for this gluing to produce a hyperbolic manifold is that S_1 and S_2 must have the same circumference. There is actually no further condition, because hyperbolic geometry is rather rigid; the geometry of a hyperbolic surface near a geodesic boundary component is completely determined by the circumference of that boundary.

Not only are we free to glue S_1 to S_2 as long as they have the same length, but an arbitrary parameter ϱ enters in this gluing, because we are free to rotate S_1 relative to S_2 . This parameter was mentioned at the end of section 3.3. We understand ϱ as the distance (not angle) of the relative rotation between S_1 and S_2 , so the range of ϱ is $0 \leq \varrho \leq \ell$, where ℓ is the circumference of S_1 or S_2 . (Alternatively, one can replace ϱ with the angle $\theta = 2\pi\varrho/\ell$, which ranges from 0 to 2π .)

The parameter ϱ can be described in the language of gauge theory and flat connections. When we glue S_1 to S_2 , we are free to make a relative gauge transformation of the fields on the two sides by any element of $\mathrm{SL}(2, \mathbb{R})$ that commutes with the monodromy around S_1 or equivalently around S_2 . If as in eqn. (3.19), the monodromy of the hyperbolic flat connection around S_1 or S_2 is by the group element $U_\ell^\pm = \pm \mathrm{diag}(e^{\ell/2}, e^{-\ell/2})$, then the element $U_\varrho^+ = \mathrm{diag}(e^{\varrho/2}, e^{-\varrho/2})$ commutes with U_ℓ , and a relative gauge transformation by U_ϱ^+ corresponds to the relative rotation between the two sides that was discussed in the last paragraph. One may wonder about the interpretation of a relative gauge transformation by the element $U_\varrho^- = -\mathrm{diag}(e^{\varrho/2}, e^{-\varrho/2})$, which also commutes with U_ℓ . The answer is that the minus sign represents a sign change of fermions on S_1 relative to those on S_2 , prior to gluing. This sign change, which produces a different spin structure on the glued manifold, was important in our discussion of the bulk dual of fermion anomalies.

Incorporating time-reversal symmetry replaces the gauge group by a double cover. In a purely bosonic theory, incorporating time-reversal symmetry means replacing $\mathrm{PSL}(2, \mathbb{R})$ by its double cover $\mathrm{PGL}(2, \mathbb{R})$ (the group of 2×2 invertible real matrices, modulo multiplication by a nonzero real scalar). The essential difference is that $\mathrm{PGL}(2, \mathbb{R})$ contains an element $\mathrm{diag}(1, -1)$ of determinant -1 , and $\mathrm{PSL}(2, \mathbb{R})$ does not. In the presence of fermions, one has to take a double cover of $\mathrm{SL}(2, \mathbb{R})$, but there are two different double covers, corresponding to pin^+ and pin^- . For pin^+ , one wants the group of 2×2 real matrices of determinant ± 1 , meaning that again one includes the element $\mathrm{diag}(1, -1)$. For pin^- , one wants the group of 2×2 matrices of determinant 1 that are either real or imaginary, meaning that one extends $\mathrm{SL}(2, \mathbb{R})$ to include the element $\mathrm{diag}(i, -i) = i \mathrm{diag}(1, -1)$. In any of these cases, one can make an orientation-reversing gluing, in which S_1 is glued to S_2 with a relative gauge transformation $\mathrm{diag}(e^{\varrho/2}, -e^{-\varrho/2})$ (or $i \mathrm{diag}(e^{\varrho/2}, -e^{-\varrho/2})$ in the pin^- case); these commute with U_ϱ^\pm . This more general type of gluing produces, for example, the twisted double trumpet that we studied in section 2. Note that in either pin^+ or pin^- , one distinguishes the group elements $\mathrm{diag}(-1, -1)$ and $\mathrm{diag}(1, 1)$, which are equivalent in $\mathrm{PGL}(2, \mathbb{R})$.

From what we have said, every gluing of two three-holed spheres is associated to a pair of moduli ℓ, ϱ . They are called length-twist or Fenchel-Nielsen coordinates. For example, if we build a genus 2 Riemann surface Y by gluing two three-holed spheres along circles S_1, S_2, S_3 (fig. 8), one gets three pairs of length-twist coordinates, making 6 parameters in all. Locally these parametrize the moduli space of Riemann surfaces of genus 2, whose real dimension is indeed 6. The subtlety is that there are many ways to build the

same Y by gluing together three-holed spheres, so the same moduli space can be parametrized locally by length-twist coordinates in many different ways. A given set of length-twist coordinates gives a good parametrization when the ℓ_i are sufficiently small, but when the ℓ_i increase beyond a certain point, there exists an alternative description by a different set of length-twist coordinates with smaller lengths. There is no simple description of the inequalities on the lengths that one should impose to get a single copy of the moduli space of hyperbolic metrics or conformal structures on Y .

3.4.2. The torsion of a three-holed sphere. In computing the torsion of a three-holed sphere, a circle, and a crosscap, we follow steps that are mostly explained more fully in sections 4.3, 4.4, and 4.6 of [90]. There are a few differences in detail because the gauge group is not compact.

We begin with a three-holed sphere and gauge group $G = \text{SL}(2, \mathbb{R})$. We henceforth write just τ rather than $\tau^{\mathbb{R}}$ for the torsion. The torsion is a measure; it will be convenient to represent this measure by a differential form. A measure is really better understood as the absolute value of a differential form, but we omit the absolute value sign. For a purely bosonic group such as $\text{SL}(2, \mathbb{R})$, the torsion is positive-definite, so we do not need to keep track of overall signs.

In general, let Y be a two-manifold and P a point in Y . Let \mathcal{R} be the moduli space of flat connections on Y , modulo gauge transformations, and $\widehat{\mathcal{R}}$ the moduli space of “based” flat connections, constructed by dividing by gauge transformations that are trivial at P . There is an obvious fibration $\pi : \widehat{\mathcal{R}} \rightarrow \mathcal{R}$, with fiber G , defined by dividing by gauge transformations that are possibly nontrivial at P .

We have already explained how the combinatorial torsion defines a measure τ on \mathcal{R} . By omitting in the definition of \mathcal{C}_0 the 0-cell that corresponds to P , but otherwise following exactly the same definitions, one can use the combinatorial torsion to define a measure $\widehat{\tau}$ on $\widehat{\mathcal{R}}$. The definition of the combinatorial torsion made use of a volume form on the Lie algebra \mathfrak{g} of G . Such a volume form also determines a volume form on the group manifold. We will call this volume form vol_G . (Note that vol_G is a volume form, not a number; we are interested in noncompact gauge groups whose volume is infinite.) The three volume forms τ , $\widehat{\tau}$, and vol_G have a simple multiplicative relationship³⁷

$$(3.24) \quad \widehat{\tau} = \text{vol}_G \cdot \pi^*(\tau).$$

³⁷This statement is equivalent to eqn. (4.39) in [90], which is formulated a little differently as the volume of G was assumed to be finite.

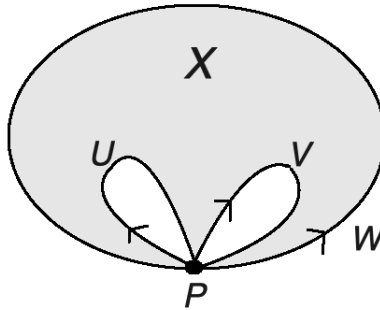


Figure 10: The shaded region is a three-holed sphere Y - or more precisely a topological space that is equivalent to Y up to homotopy. It has a simple cell decomposition with a single 0-cell P , three 1-cells that have been labeled by the holonomies U, V, W of a flat connection, and a single two-cell X .

In general, omitting a single 0-cell in the definition of \mathcal{C}_0 might not lead to a major simplification. However, for the three-holed sphere (or more exactly for a topological space that is equivalent to a three-holed sphere up to homotopy, which is good enough in computing the torsion), one can pick a very simple cell decomposition with only a single 0-cell P (fig. 10). This means that after omitting P , we get $\mathcal{C}_0 = 0$, and this does lead to a major simplification. In the cell decomposition of the figure, there is a single 2-cell X , and three 1-cells that have been labeled by the holonomies U, V, W of a flat connection. Hence $\mathcal{C}_2 = \mathcal{V}_X$, $\mathcal{C}_1 = \mathcal{V}_U \oplus \mathcal{V}_V \oplus \mathcal{V}_W$. Since $\mathcal{C}_0 = 0$, we can compute the torsion just using the boundary map $\partial : \mathcal{C}_2 \rightarrow \mathcal{C}_1$. But matters are even simpler than that. If $\Theta : \mathcal{C}_1 \rightarrow \mathcal{V}_W$ is the obvious projection, then $\Theta \circ \partial : \mathcal{C}_2 \rightarrow \mathcal{V}_W$ is an isomorphism. It follows then from the definition of the torsion that if we discard \mathcal{C}_2 and \mathcal{V}_W , the torsion is unchanged. This means that we can compute the torsion $\hat{\tau}$ taking $\mathcal{C}_2 = \mathcal{C}_0 = \partial = 0$, $\mathcal{C}_1 = \mathcal{V}_U \oplus \mathcal{V}_V$.

On a three-holed sphere, the holonomies of a flat connection are group elements U, V, W satisfying $UVW = 1$. To construct the moduli space \mathcal{R} , we would identify two such triples that differ by conjugation by an element R of G : $(U, V, W) \cong (RUR^{-1}, RVR^{-1}, RWR^{-1})$. To construct the based moduli space $\hat{\mathcal{R}}$, we do not divide by conjugation. Since $UVW = 1$, $\hat{\mathcal{R}} \cong G \times G$, parametrized by U and V . With $\mathcal{C}_2 = \mathcal{C}_0 = \partial = 0$, $\mathcal{C}_1 = \mathcal{V}_U \oplus \mathcal{V}_V$, eqn. (3.15) just says that $\hat{\tau}$ is the natural volume form on $G \times G$. Since there are several copies of G in this discussion, we will write $\text{vol}_G(U)$ for the volume form of

a copy of G parametrized by U , and similarly for $\text{vol}_G(V)$ and $\text{vol}_G(R)$. So

$$(3.25) \quad \widehat{\tau} = \text{vol}_G(U) \cdot \text{vol}_G(V).$$

We parametrize U and V by the obvious R -dependent generalization of eqn. (3.21):

$$(3.26) \quad U = RU_0R^{-1}, \quad V = RV_0R^{-1},$$

with

$$(3.27) \quad U_0 = \begin{pmatrix} e^{a/2} & \kappa \\ 0 & e^{-a/2} \end{pmatrix}, \quad V_0 = \begin{pmatrix} e^{-b/2} & 0 \\ 1 & e^{b/2} \end{pmatrix}.$$

Thus a, b , and κ (or U_0 and V_0) parametrize the ordinary moduli space \mathcal{R} , while a, b, κ , and R parametrize the extended moduli space $\widehat{\mathcal{R}}$. Eqns. (3.24) and (3.25) tell us that

$$(3.28) \quad \text{vol}_G(U) \cdot \text{vol}_G(V) = \text{vol}_G(R) \cdot \tau,$$

and this condition will determine τ .

We introduce a standard basis for the Lie algebra \mathfrak{sl}_2 of $\text{SL}(2, \mathbb{R})$:

$$(3.29) \quad e = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad f = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \quad h = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Now we can describe explicitly what measures we will use on the Lie algebra and on the group. If x is an element of the Lie algebra \mathfrak{sl}_2 , write it as $x = \begin{pmatrix} x_h & x_e \\ x_f & -x_h \end{pmatrix} = x_e e + x_f f + x_h h$, and then define the measure on \mathfrak{sl}_2 to be $4dx_e dx_f dx_h$.³⁸ Similarly, if U is an element of the group $\text{SL}(2, \mathbb{R})$, then $U^{-1}dU$ is a Lie algebra valued one-form, and we expand it as $U^{-1}dU =$

³⁸ The factor of 4 has been chosen with foresight so that our final result for a measure on moduli space will agree (in the oriented case) with standard formulas. The measure we are using can be derived from the inner product $(x, x) = 2 \text{Tr } x^2 = 4x_h^2 + 4x_e x_f$ on the Lie algebra. This inner product corresponds to the metric $ds^2 = 4dx_h^2 + 4dx_e dx_f$ on \mathfrak{g} , so that the metric tensor g satisfies $\sqrt{|\det g|} = 4$ and the Riemannian measure is $4dx_e dx_f dx_h$. Using a different measure would multiply the final result for the torsion by a factor $(\text{const.})^\chi$ where χ is the Euler characteristic. Such a factor can be absorbed by adjusting the coefficient of χ in the action of the underlying JT theory.

$(U^{-1}dU)_e e + (U^{-1}dU)_f f + (U^{-1}dU)_h h$. Then the volume form on the group is the three-form

$$(3.30) \quad \text{vol}_G(U) = 4(U^{-1}dU)_e (U^{-1}dU)_f (U^{-1}dU)_h.$$

To use equation (3.28) to compute τ , it suffices to work at $R = 1$. Expanding $R = 1 + r + \mathcal{O}(r^2)$, we have $dR = dr + \mathcal{O}(r)$, $dR^{-1} = -dr + \mathcal{O}(r)$. Thus at $R = 1$,

$$\text{vol}_G(R) = 4dr_e dr_f dr_h.$$

The general form of τ is $F(a, b, \kappa) da db d\kappa$, for some function $F(a, b, \kappa)$ that we want to compute. Thus the right hand side of eqn. (3.28) reads

$$(3.31) \quad 4dr_e dr_f dr_h F(a, b, \kappa) da db d\kappa,$$

and we need to compare this to the left hand side.

To evaluate the left hand side, we first note that at $R = 1$,

$$(3.32) \quad \begin{aligned} U^{-1}dU &= U_0^{-1}dU_0 + U_0^{-1}drU_0 - dr, \\ V^{-1}dV &= V_0^{-1}dV_0 + V_0^{-1}drV_0 - dr. \end{aligned}$$

Since U_0 is upper triangular, there is a simple result for the lower triangular part of $U^{-1}dU$:

$$(3.33) \quad (U^{-1}dU)_f = (e^a - 1)dr_f.$$

Similarly, there is a simple result for the upper triangular part of $V^{-1}dV$:

$$(3.34) \quad (V^{-1}dV)_e = (e^b - 1)dr_e.$$

We can now eliminate dr_f , $(U^{-1}dU)_f$, dr_e , and $(V^{-1}dV)_e$ from eqn. (3.28), reducing to

$$(3.35) \quad \begin{aligned} &4(e^a - 1)(e^b - 1)(U^{-1}dU)_e (U^{-1}dU)_h (V^{-1}dV)_f (V^{-1}dV)_h \\ &= dr_h F(a, b, \kappa) da db d\kappa. \end{aligned}$$

In the further evaluation of the left hand side of eqn. (3.35), we can set $dr_e = dr_f = 0$. After doing so, we get some simple formulas:

$$\begin{aligned}
 (U^{-1}dU)_h &= \frac{1}{2}da \\
 (U^{-1}dU)_e &= e^{-a/2}d\kappa + \frac{\kappa}{2}e^{-a/2}da + 2\kappa e^{-a/2}dr_h \\
 (V^{-1}dV)_h &= -\frac{1}{2}db \\
 (3.36) \quad (V^{-1}dV)_f &= \frac{1}{2}e^{-b/2}db - 2e^{-b/2}dr_h.
 \end{aligned}$$

The left hand side of eqn. (3.35) is therefore just $8 \sinh \frac{a}{2} \sinh \frac{b}{2} dr_h da db d\kappa$. From eqn. (3.35), we therefore get

$$(3.37) \quad \tau = F(a, b, \kappa)da db d\kappa = 8 \sinh \frac{a}{2} \sinh \frac{b}{2} da db d\kappa.$$

Concretely, starting with eqn. (3.35), we first replace $(U^{-1}dU)_h (V^{-1}dV)_h$ in (3.35) with $-\frac{1}{4}da db$, cancel da and db from the equation, and set $da = db = 0$ in the formulas for $(U^{-1}dU)_e$ and $(V^{-1}dV)_f$. Then $(U^{-1}dU)_e(V^{-1}dV)_f$ reduces to $-2e^{-a/2-b/2}d\kappa dr_h$, and this leads to eqn. (3.37).

In view of eqn. (3.23), we can here replace $d\kappa$ with $\sinh \frac{c}{2} dc$, where c is the length parameter of the third boundary component. Thus finally the torsion of a three-holed sphere Y with boundary length parameters a, b, c is

$$(3.38) \quad \tau_Y = 8 \sinh \frac{a}{2} \sinh \frac{b}{2} \sinh \frac{c}{2} da db dc.$$

As a check on the calculation, we note that this is symmetric in a, b, c , though the derivation lacked that symmetry. We recall that the overall sign in eqn. (3.38) is unimportant, as the torsion of the three-holed sphere is really the absolute value of the three-form that appears on the right hand side. The factor of 8 is meaningful but depended on how we normalized the measure on the Lie algebra.

3.4.3. The torsion of a circle. Now we consider a circle S endowed with a flat $SL(2, \mathbb{R})$ bundle $E \rightarrow S$. We take the holonomy of this flat bundle to be $U = \text{diag}(e^{a/2}, e^{-a/2})$.

We can endow S with a very simple cell decomposition, with a single 0-cell P and a single 1-cell C (fig. 11). The computation of the torsion is thus going to be rather simple. \mathcal{C}_1 and \mathcal{C}_0 are both copies of the Lie algebra \mathfrak{sl}_2 , while $\mathcal{C}_2 = 0$. In the definition of the boundary operator $\partial : \mathcal{V}_1 \rightarrow \mathcal{V}_0$, we

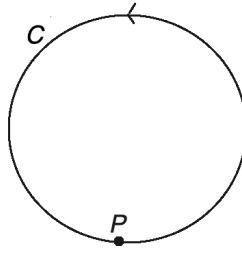


Figure 11: A simple cell decomposition of a circle, with one 0-cell P , and one 1-cell C , with holonomy U .

need to take into account that both ends of C are on P . Let us trivialize the bundle E at and just to the right of the point P ; we use this trivialization in identifying \mathcal{V}_1 and \mathcal{V}_0 with \mathfrak{sl}_2 . For $s \in \mathcal{V}_1$, the boundary operator acts on S by

$$(3.39) \quad \partial s = UsU^{-1} - s.$$

The two terms are the contributions of the two ends of C . The contribution from C ending on P from the left is proportional to UsU^{-1} , because to evaluate it we have to parallel transport s along C from one end of C to the other. There is a relative minus sign between the two terms because of a relative orientation between the two ends of C .

It is convenient to decompose the \mathfrak{sl}_2 Lie algebra as $\mathfrak{sl}_2^0 \oplus \mathfrak{sl}_2^\perp$, where \mathfrak{sl}_2^0 consists of diagonal matrices and \mathfrak{sl}_2^\perp consists of matrices whose diagonal matrix elements vanish. The operator ∂ commutes with this decomposition, so $\partial = \partial^0 \oplus \partial^\perp$, where ∂^0 and ∂^\perp act on \mathfrak{sl}_2^0 and on \mathfrak{sl}_2^\perp , respectively. The torsion τ_S of a flat bundle on S correspondingly factorizes

$$(3.40) \quad \tau_S = \tau_S^\perp \cdot \tau_S^0,$$

where τ_S^\perp and τ_S^0 are the torsion defined using ∂^\perp and ∂^0 , respectively.

The operator $\partial^\perp : \mathfrak{sl}_2^\perp \rightarrow \mathfrak{sl}_2^\perp$ is invertible, so we can use the simplest definition (3.14) of the torsion. In the present case, with $\mathcal{V}_2 = 0$, and $\mathcal{V}_1, \mathcal{V}_0$ both copies of \mathfrak{sl}_2^\perp , the definition (3.14) reduces to $\tau^\perp = \alpha_0(\partial^\perp t_1, \partial^\perp t_2) / \alpha_1(t_1, t_2)$, where t_1, t_2 is any basis of \mathfrak{sl}_2^\perp . But with the two measures α_0 and α_1 being the same, this ratio is, according to eqn. (3.13), the absolute value of the determinant $|\det \partial^\perp|$. The eigenvalues of ∂^\perp are $e^{\pm a} - 1$, so

$$(3.41) \quad \tau_S^\perp = |\det \partial^\perp| = |(e^a - 1)(e^{-a} - 1)| = 4 \sinh^2 \frac{a}{2}.$$

On the other hand, ∂^0 is actually identically 0, since $UsU^{-1} - s = 0$ if s is diagonal. This being so, the definition (3.17) reduces to $\tau^0 = \alpha_0(v)/\alpha_1(u)$, where v and u are basis vectors of the 1-dimensional spaces \mathcal{C}_0^0 and \mathcal{C}_1^0 . This just means that τ^0 is the ratio of the natural measures on \mathcal{C}_0^0 and \mathcal{C}_1^0 . Since those spaces are both copies of \mathfrak{sl}_2^0 , the natural measures are the same, and one could claim that $\tau^0 = 1$. It is more convenient not to take that step because, although \mathcal{C}_0 and \mathcal{C}_1 are both copies of \mathfrak{sl}_2 , the physical meaning is different. \mathcal{C}_1 is the tangent space to the moduli space of flat connections on the circle. \mathcal{C}_0 is the Lie algebra of the subgroup of $SL(2, \mathbb{R})$ that commutes with a given flat connection. If we parametrize the moduli space by the parameter a that appears in the monodromy $U = \text{diag}(e^{a/2}, e^{-a/2})$, and similarly write $\text{diag}(e^{\varrho/2}, e^{-\varrho/2})$ for a group element that commutes with U , then the ratio of the natural measures is

$$(3.42) \quad \tau_S^0 = da \cdot (d\varrho)^{-1}.$$

Combining these statements, we get the torsion of a flat connection on the circle:

$$(3.43) \quad \tau_S = 4 \sinh^2 \frac{a}{2} \cdot da \cdot (d\varrho)^{-1}.$$

3.4.4. The torsion of an oriented surface. Let Y_1 and Y_2 be two three-holed spheres endowed with hyperbolic metrics, and with respective boundary components S_1 and S_2 , each with length parameter a . As explained in section 3.4.1, we can build a larger hyperbolic manifold Y by gluing $S_1 \subset Y_1$ onto $S_2 \subset Y_2$. According to eqn. (3.18), the torsion of Y is related to the torsions of Y_1 , Y_2 , and $S = S_1 = S_2$ by $\tau_Y = \tau_{Y_1} \tau_S^{-1} \tau_{Y_2}$.

Let us isolate in this formula the factors that depend on variables that are defined along S . From both τ_{Y_1} and τ_{Y_2} , we get a factor of $2 \sinh \frac{a}{2} da$. On the other hand, from τ_S^{-1} , we get $d\varrho/4 \sinh^2 \frac{a}{2} da$. When we multiply these quantities together, the factors of $\sinh a/2$ cancel, and we are left with just $dad\varrho$. As promised in section 3.3, an unwanted extra factor of da has disappeared. We have also gotten a factor of $d\varrho$, a measure on the subgroup of $SL(2, \mathbb{R})$ that commutes with the monodromy around S . We can, in other words, think of $d\varrho$ as a measure for integrating over the gluing parameter ϱ that was described in section 3.3.

Since any oriented two-manifold can be constructed by gluing together three-holed spheres, we can, without further ado, construct the torsion for an arbitrary oriented two-manifold. For simplicity, let us consider the case of an oriented two-manifold Y of genus g without boundary. It can be built

by gluing together $2g - 2$ three-holed spheres Y_α along $3g - 3$ circles S_i (for $g = 2$, this was drawn in fig. 8). For each circle, we get a pair of length-twist coordinates a_i and ϱ_i , and the above computation leads to an integration measure $da_i d\varrho_i$ for those variables. We also get a factor 8 for each three-holed sphere, which cancels the $1/4$ for each circle. Thus the volume form for the moduli space \mathcal{M} of hyperbolic structures on Y that we get from the torsion is

$$(3.44) \quad \tau_Y = \prod_{i=1}^{3g-3} da_i d\varrho_i.$$

For a different normalization of the measure on \mathfrak{sl}_2 , this would be multiplied by $(\text{const.})^X$.

A few words about this formula are in order. First, length-twist coordinates on \mathcal{M} are far from unique, since there are infinitely many ways to decompose Y into a union of three-holed spheres. The formula (3.44) is valid for any choice of length-twist coordinates.

Second, this formula for the measure (for the case of an oriented two-manifold Y) is actually well-known, but it is not usually derived using the torsion. A more common procedure is to use the fact that if Y is oriented, then \mathcal{M} is a symplectic (and in fact Kahler) manifold. The symplectic form of \mathcal{M} in length-twist coordinates is $\sum_i da_i d\varrho_i$, and this leads to the volume form (3.44). The two approaches give the same volume form because, as we described in the beginning of this section, the torsion is “trivial” for orientable Y , which just means that it is equivalent to the volume form that can be defined using the symplectic or Riemannian structure of \mathcal{M} . The symplectic or Riemannian structure of \mathcal{M} is more elementary than the torsion in the sense that these structures are defined just in terms of zero-mode wavefunctions on Y , while the torsion involves a quantum correction coming from the non-zero modes. There are two reasons that we have used the torsion to arrive at the formula for the volume form: this is the natural definition in JT gravity, and (therefore) it generalizes to unorientable two-manifolds.

Finally, though the formula (3.44) is simple, it does not lead to any simple results for the volume of \mathcal{M} , because the region in length-twist coordinates that corresponds to \mathcal{M} is not simple to describe. This is actually the problem that was overcome in the work of Maryam Mirzakhani on the volumes of moduli spaces.

3.4.5. The torsion of a cross-cap. To complete the picture, we also need to compute the torsion of a crosscap.

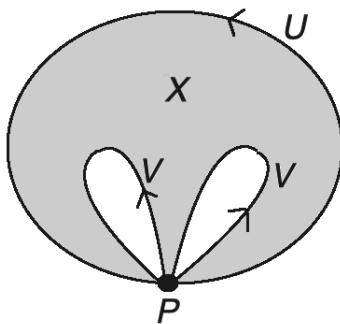


Figure 12: The shaded region in this picture is the crosscap manifold C or more exactly a topological space equivalent to it in homotopy. It has a simple cell decomposition with one 2-cell X , two 1-cells labeled by their holonomies U and V , and one 0-cell P . V appears twice in the drawing, but the two copies are supposed to be identified to create C .

By a crosscap, we mean here a two-manifold C that is constructed from a cylinder by making a suitable identification on one boundary. For example, we can start with the cylinder $I \times S^1$, where I is the unit interval $0 \leq x \leq 1$, and S^1 is the circle $\theta \cong \theta + 2\pi$. Then at $x = 1$, we make the identification $\theta \cong \theta + \pi$, while doing nothing at $x = 0$. This gives an unorientable two-manifold C , whose boundary is a single circle S , the circle at $x = 0$. At $x = 1$, there is a second circle S' with $0 \leq \theta \leq \pi$, but because of the gluing, this is an “internal” circle, not a boundary. C can be viewed as a Möbius strip or as a copy of \mathbb{RP}^2 with an open ball removed.

Let U be the monodromy of a flat connection on C around S , and let V be the monodromy of the same connection around S' . Clearly they are related by $U = V^2$. For hyperbolic monodromies, up to conjugacy we can take $U = \text{diag}(e^{a/2}, e^{-a/2})$, $V = \text{diag}(e^{a/4}, -e^{-a/4})$. There is an important minus sign in the formula for V which reflects the fact that the orientation of C is reversed in going around S' . Spin structures are not important in the discussion of the torsion (because the fields of the BF model are all in the adjoint representation of $\text{SL}(2, \mathbb{R})$) so for simplicity we ignore the overall sign of the monodromies and subtleties of pin^- and pin^+ . (We will have to incorporate those subtleties when we get to the supersymmetric case.)

C has a simple cell decomposition (fig. 12), with one 2-cell X , a pair of 1-cells associated to holonomies U and V , respectively, and one 0-cell P . Thus $\mathcal{C}_2 \cong \mathfrak{sl}_2$, $\mathcal{C}_1 \cong \mathfrak{sl}_2 \oplus \mathfrak{sl}_2$, with the two summands corresponding to the 1-cells labeled U and V , respectively, and $\mathcal{C}_0 \cong \mathfrak{sl}_2$. In defining the boundary

map, we trivialize the flat bundle E just to the right of the point P in the figure. Then the boundary map $\partial_2 : \mathcal{C}_2 \rightarrow \mathcal{C}_1$ is

$$(3.45) \quad \partial_2(s) = s \oplus (-s - VsV^{-1}),$$

while $\partial_1 : \mathcal{C}_1 \rightarrow \mathcal{C}_0$ is

$$(3.46) \quad \partial_1(t \oplus u) = UtU^{-1} - t + VuV^{-1} - u.$$

One may verify that $\partial_1 \circ \partial_2 = 0$, using $V^2 = U$.

Because V and U are diagonal, the torsion of C has the same sort of factorization as in section 3.4.3:

$$(3.47) \quad \tau_C = \tau_C^\perp \cdot \tau_C^0.$$

To compute τ_C^\perp and τ_C^0 , note first that if we compose ∂_2 with the projection map on the first summand of $\mathcal{C}_1 = \mathfrak{sl}_2 \oplus \mathfrak{sl}_2$, it is the identity. This means that τ_C^\perp is unchanged if we just drop \mathcal{C}_2 and the first summand of \mathcal{C}_1 . This reduces us to the torsion of the map $\partial'_1 : \mathfrak{sl}_2 \rightarrow \mathfrak{sl}_2$ defined by $\partial'_1(u) = VuV^{-1} - u$.

From here on, the discussion is similar to the discussion of the torsion of a circle in section 3.4.3. Restricted to \mathfrak{sl}_2^\perp , ∂'_1 is invertible, and τ_C^\perp is just the absolute value of its determinant. The eigenvalues of ∂'_1 acting on \mathfrak{sl}_2^\perp are $-1 - e^{\pm a/2}$, so its determinant is $(1 + e^{a/2})(1 + e^{-a/2}) = 4 \cosh^2 a/4$:

$$(3.48) \quad \tau_C^\perp = 4 \cosh^2 \frac{a}{4}.$$

On the other hand, restricted to \mathfrak{sl}_2^0 , $\partial'_1 = 0$. So its torsion is just a ratio of two copies of the natural measure on \mathfrak{sl}_2^0 . As in section 3.4.3, though we could call this ratio 1, it is more useful to think of it as a ratio of measures on two copies of \mathfrak{sl}_2^0 : one copy is the tangent bundle to the moduli space of flat connections on C , and the other copy generates the automorphism group of such a flat connection. Relative to section 3.4.3, we have to be careful with a factor of 2. The definition of ∂'_1 is such that ∂'_1 maps flat sections on S' (elements of the second summand in \mathcal{C}_1) to flat sections at p , but when we glue C to some other Riemann surface, the gluing takes place along S , not S' . So what was $da \cdot d\rho^{-1}$ in section 3.4.3 here becomes $d\bar{a} \cdot d\rho^{-1}$, where \bar{a} is the length parameter of S' , but ρ generates twists of S . Since $\bar{a} = a/2$ (the monodromy around S is $\text{diag}(e^{a/2}, e^{-a/2})$, while that

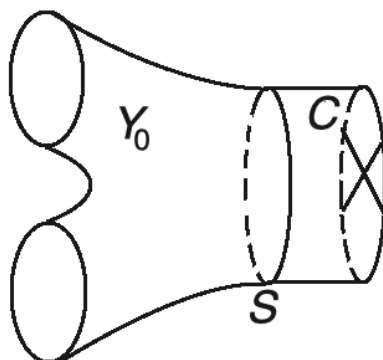


Figure 13: A boundary of a two-manifold Y_0 can be replaced by a crosscap by gluing the crosscap manifold C to a boundary S of Y_0 .

around S' is $\text{diag}(e^{\bar{a}/2}, -e^{-\bar{a}/2})$, we have

$$(3.49) \quad \tau_C^0 = d\bar{a} \cdot (d\varrho)^{-1} = \frac{1}{2} da \cdot (d\varrho)^{-1}$$

and

$$(3.50) \quad \tau_C = 2 \cosh^2 \frac{a}{4} \cdot da \cdot (d\varrho)^{-1}.$$

3.4.6. The torsion of an unorientable two-manifold. Now let Y be an unorientable two-manifold, without boundary for simplicity. Such a manifold is not a complex Riemann surface, but – assuming its Euler characteristic is negative – it does admit a hyperbolic metric, that is a metric of constant scalar curvature $R = -2$.

A compact oriented hyperbolic two-manifold can be built by gluing together three-holed hyperbolic spheres with geodesic boundaries. To build a hyperbolic metric on an arbitrary possibly unorientable Y , we need another kind of building block. This is obtained from a three-holed sphere Y_0 by replacing one or more of its boundary circles with crosscaps. If one of the boundaries of Y_0 is a geodesic circle S defined by $\theta \cong \theta + 2\pi$, then to replace S with a crosscap, we just make a further identification $\theta \cong \theta + \pi$, after which Y_0 becomes unorientable and S is replaced by a circle S' that is a geodesic of one-half the length of S . S' is a “one-sided” geodesic, because in going around S' , its two sides are exchanged.

If we make such a replacement on one or two boundary components of Y_0 , we get a hyperbolic two-manifold Y' with two or one remaining boundary components that can be used in further gluing.³⁹ In such gluing, we use as “elementary building blocks” either a three-holed sphere or its modification by replacing some boundaries with crosscaps, as just described. By gluing such elementary building blocks together along their geodesic boundaries, one can construct (in general in infinitely many different ways) hyperbolic metrics on an arbitrary possibly unorientable two-manifold Y without boundary.⁴⁰

If Y is presented in such a fashion, then the moduli space of hyperbolic structures on Y can be described locally by a generalization of the length-twist or Fenchel-Nielsen coordinates that we used in section 3.4.4. Gluing of two elementary building blocks along a boundary circle always produces a pair of length-twist moduli a, ϱ . But if one of the elementary building blocks has a one-sided geodesic as one of its ends, then this one-sided geodesic carries a length parameter but no corresponding twist parameter, as it is not going to be glued onto anything.

To get a general formula for the torsion of any two-manifold in length-twist coordinates, we just need to know what happens when a boundary circle is replaced by a one-sided geodesic. There is a local universal formula for this. Let Y_0 be a three-holed sphere and let C be the crosscap manifold whose torsion was computed in section 3.4.5. Thus C has a single boundary circle S . By gluing S onto one of the boundary circles of Y_0 , we make a new manifold Y' in which a boundary circle has been replaced by a one-sided geodesic (fig. 13). Since we already have computed the torsions of Y_0 , C , and S , it is straightforward to compute the torsion of Y' using the gluing law:

$$(3.51) \quad \tau_{Y'} = \tau_{Y_0} \frac{1}{\tau_S} \tau_C.$$

We focus on the factors in the torsion that involve the boundary component S of Y_0 at which the gluing is occurring. If the holonomy around S

³⁹If we make the same replacement on all three boundary components of Y_0 , we get a hyperbolic structure on a particular compact two-manifold Y' without boundary. As Y' has no boundary, it is not a building block for further gluing. But the discussion in the text applies for computing its torsion.

⁴⁰The gluing maps used here can be orientation-reversing. If Y is unorientable but has even (negative) Euler characteristic, then hyperbolic metrics on Y can be constructed without the use of crosscaps by gluing together three-holed spheres with gluing maps some of which are orientation-reversing.

is $\text{diag}(e^{a/2}, e^{-a/2})$ (and therefore the length parameter of the one-sided geodesic in C is $\bar{a} = a/2$), then the factor in τ_{Y_0} that involves a is $2 \sinh \frac{a}{2} da$. Including additional a -dependent factors in eqns. (3.43) and (3.50), one finds that the a -dependent part of $\tau_{Y_0} \tau_S^{-1} \tau_C$ is $\frac{1}{2} \coth \frac{a}{4} da$.

A two-manifold Y of genus g with n crosscaps attached can be built by gluing together $2g - 2 + n$ building blocks, namely three-holed spheres with a total of n boundaries replaced by crosscaps, along $3g - 3 + n$ circles. The Euler characteristic of such a surface is $\chi = 2 - 2g - n$. Associated with the i^{th} circle is a pair of length-twist parameters a_i, ϱ_i , and associated with the α^{th} crosscap is a length parameter $a_\alpha/2$ (the length of its one-sided geodesic). Putting the pieces together, and keeping track of factors of 2, we learn that the torsion of Y is

$$(3.52) \quad \tau_Y = \prod_{i=1}^{3g-3} da_i d\varrho_i \prod_{\alpha=1}^n \frac{1}{2} \coth \frac{a_\alpha}{4} da_\alpha.$$

This volume form on the moduli space of hyperbolic structures on an unorientable two-manifold Y was actually first defined by Norbury (see Theorem 5 in [52]; note that Norbury’s ℓ_α is our $a_\alpha/2$). This volume form has been further studied in [93]. The starting point in [52] was not the torsion. Rather, the starting point was the hypothesis that there might be a volume form that takes a simple factorized form in length-twist coordinates and is independent of the particular choice of such coordinates. As there are infinitely many sets of length-twist coordinates on a given surface, the condition that a particular factorized expression does not depend on the choice is very restrictive. It turned out that the expression in eqn. (3.52) does have this property (and no other factorized expression does [93]). In our derivation here, the starting point was the torsion, which manifestly is a topological invariant, not depending on a choice of length-twist coordinates. That starting point did not make it completely obvious that we would get a factorized expression in length-twist coordinates, but the gluing law for the torsion goes a long way towards explaining why this happens.

As originally noted in [52] and further elaborated in [93], if one actually tries to integrate the volume form τ_Y to compute the volume of the moduli space, one runs into what in string theory would be called an infrared divergence. Indeed, the measure behaves as da_α/a_α for small a_α .

3.4.7. Application to JT gravity. In applications to JT gravity, one is usually not interested in a compact two-manifold Y , but in a two-manifold with non-compact “ends” that are asymptotic to a cutoff version of AdS_2 ,

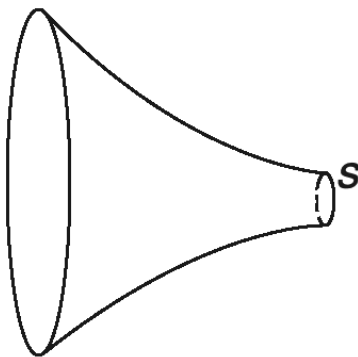


Figure 14: A trumpet.

which we will call NAdS_2 . A Schwarzian mode propagates along the boundary of NAdS_2 .

The most basic example is NAdS_2 itself. An important ingredient in describing other possibilities is a “trumpet” T (fig. 14), which looks like NAdS_2 on the “outside” and has an “internal” boundary consisting of a circle S whose boundary is a geodesic with length parameter a . The boundary of T can be glued onto a boundary component of a compact two-manifold, or onto a two-manifold with additional noncompact ends. For example, one can glue together two trumpets along a common boundary of length a to make a “double trumpet,” which we already discussed in section 2.4.1. Let $Z_{\text{JT}}^T(a)$ be the JT path integral on a trumpet T whose internal boundary has length a . An important factor in $Z_{\text{JT}}^T(a)$ is the path integral of the Schwarzian mode. The JT path integral of a double trumpet is the product of two factors of $Z_{\text{JT}}^T(a)$, one for each trumpet, integrated over a and over the gluing parameter ρ . The measure that must be integrated is

$$(3.53) \quad Z_{\text{JT}}^T(a) \, d a d \rho \, Z_{\text{JT}}^T(a).$$

It is not immediately clear how to compute $Z_{\text{JT}}^T(a)$ in the language of the torsion. It may be hard to combine the combinatorial approach to the torsion with the boundary conditions that are needed to include the Schwarzian mode on the “outer” boundary of the trumpet. Even on the “inner” boundary of the trumpet, there is a nontrivial issue to consider: the boundary condition that we have used in computing the torsion on a manifold with boundary actually does not coincide with what is usually used in defining the JT trumpet path integral $Z_{\text{JT}}^T(a)$. That is why in the gluing law for

the torsion, which we described in section 3.3, one has to divide by τ_S , the torsion of a circle.

Rather than trying to deal with these issues directly, we will take a shortcut. We will write $\tilde{Z}_{\text{JT}}^T(a)da$ for a JT path integral on the trumpet computed with “Schwarzian” boundary conditions on the outer boundary, and with “torsion” boundary conditions on the inner boundary. We will not try to compute $\tilde{Z}_{\text{JT}}^T(a)$ directly, but rather we will infer it by comparing to eqn. (3.53). To get from $\tilde{Z}_{\text{JT}}^T(a)da$ to a JT path integral on the double trumpet, we simply glue together two trumpets on their inner boundaries, using the torsion gluing rule since $\tilde{Z}_{\text{JT}}^T(a)da$ is defined with “torsion” boundary conditions on the inner boundary. So eqn. (3.53) should be compared to $\tilde{Z}_{\text{JT}}^T(a)da\tau_S^{-1}\tilde{Z}_{\text{JT}}^T(a)da$. Using eqn. (3.43) for τ_S , we get⁴¹

$$(3.54) \quad Z_{\text{JT}}^T(a)dad\varrho Z_{\text{JT}}^T(a) = \tilde{Z}_{\text{JT}}^T(a) \frac{dad\varrho}{4 \sinh^2 \frac{a}{2}} \tilde{Z}_{\text{JT}}^T(a).$$

So

$$(3.55) \quad \tilde{Z}_{\text{JT}}^T(a) = Z_{\text{JT}}^T \cdot 2 \sinh \frac{a}{2}.$$

(Since the bosonic JT path integral is positive and the torsion is a positive measure, we do not need to worry about a sign here in extracting a square root.)

Now that we know the trumpet path integral with “torsion” boundary conditions on the inner boundary, we can predict the path integral for any manifold made by gluing a compact surface Y onto a trumpet (or a collection of trumpets). We just treat Y and its gluing onto the trumpet using the usual rules for the torsion. For example, the basic crosscap spacetime is made by replacing the boundary of the trumpet with a crosscap. Replacing a boundary with a crosscap is done by gluing in a copy of the crosscap manifold C , so it gives a factor of $\frac{1}{\tau_S}\tau_C$, where τ_S is given in eqn. (3.43) and τ_C in eqn. (3.50). Using also eqn. (3.55) for $\tilde{Z}_{\text{JT}}^T(a)$, we find that the path integral measure for the crosscap spacetime is simply

$$(3.56) \quad Z_{\text{JT}}^T \cdot \frac{1}{2} \coth \frac{a}{4} da,$$

where the circumference of the one-sided geodesic in the crosscap is $\bar{a} = a/2$.

⁴¹Recall that we are really discussing measures, so the overall signs in these formulas are not important.

In a theory with fermions, this is the result for a particular pin^- or pin^+ structure. It must be multiplied by another factor that arises from the sum over pin^\pm structures, as described in section 2.5.3.

3.5. Supersymmetry and the torsion

3.5.1. Supersymmetry and time-reversal in two dimensions. As discussed in section 2.5.1, a time-reversal symmetry \mathbb{T} , in a theory with no relevant global symmetry other than $(-1)^F$, will satisfy either $\mathbb{T}^2 = 1$ or $\mathbb{T}^2 = (-1)^F$. The two cases correspond to pin^- and pin^+ , respectively. In either case, there might be a c -number anomaly, so the distinction is really that for pin^- , \mathbb{T}^2 commutes with fermionic fields, and for pin^+ , it anticommutes with them.

In a quantum mechanical system with just a single supercharge Q , a \mathbb{T} symmetry is necessarily of pin^- type, since $\mathbb{T}^2 = (-1)^F$ would force fermionic operators (such as the supercharge Q) to come in pairs.

A similar statement holds in a bulk dual theory, when there is one. As we will explain, in a two-dimensional model with $\mathcal{N} = 1$ supersymmetry and no relevant global symmetries, time-reversal symmetry of pin^- type is the only option. Explaining this will involve a few details, but the necessary details will anyway serve as preparation for analyzing the torsion in $\mathcal{N} = 1$ supergravity.

In general, in D dimensions, AdS_D spacetime has symmetry $\text{SO}(2, D - 1)$ in Lorentz signature, or $\text{SO}(1, D)$ in Euclidean signature. $D = 2$ happens to be the only case in which these groups are isomorphic. We start in Lorentz signature with AdS_2 defined as the universal cover of the space parametrized by X, Y, Z with a constraint

$$(3.57) \quad X^2 + Y^2 - Z^2 = R^2,$$

where R is the radius of curvature, and the metric tensor is $ds^2 = -dX^2 - dY^2 + dZ^2$. The constraint (3.57) can be solved by introducing a time coordinate t with

$$(3.58) \quad X = \sqrt{R^2 + Z^2} \cos t, \quad Y = \sqrt{R^2 + Z^2} \sin t.$$

This makes it evident that we can take time-reversal \mathbb{T} to act by $(X, Y, Z) \rightarrow (X, -Y, Z)$, while $(X, Y, Z) \rightarrow (X, Y, -Z)$ is a spatial reflection \mathbb{R} . In a relativistic theory, \mathbb{T} is a symmetry if and only if \mathbb{R} is, and (in the absence of other symmetries with which \mathbb{T} or \mathbb{R} could be combined) the properties of \mathbb{T}

determine the properties of R . We will consider both T and R because this will make the passage to Euclidean signature obvious.

To include spin, but not supersymmetry, one replaces $SO(2, 1)$ by its double cover $SL(2, \mathbb{R})$. In the generalization with $\mathcal{N} = 1$ supersymmetry, $SL(2, \mathbb{R})$ is extended to the supergroup $OSp(1|2)$, which can be regarded as the group of linear transformations of two bosonic variables u, v and one fermionic variable θ that preserve the symplectic form

$$(3.59) \quad \widehat{\omega} = dudv + \frac{1}{2}d\theta^2.$$

The corresponding Lie superalgebra $osp(1|2)$ can be described explicitly in terms of matrices acting on $u, v|\theta$. It will be helpful to have a concrete description of the Lie superalgebra $osp(1|2)$. This Lie algebra has dimension $3|2$ – three bosonic generators and two fermionic ones. Acting on the triple

$\begin{pmatrix} u \\ v \\ \theta \end{pmatrix}$, one can pick a basis of bosonic generators

$$(3.60) \quad \begin{aligned} e &= \left(\begin{array}{cc|c} 0 & 1 & 0 \\ 0 & 0 & 0 \\ \hline 0 & 0 & 0 \end{array} \right), & f &= \left(\begin{array}{cc|c} 0 & 0 & 0 \\ 1 & 0 & 0 \\ \hline 0 & 0 & 0 \end{array} \right), \\ h &= \left(\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & -1 & 0 \\ \hline 0 & 0 & 0 \end{array} \right) \end{aligned}$$

and fermionic ones

$$(3.61) \quad q_1 = \left(\begin{array}{cc|c} 0 & 0 & 1 \\ 0 & 0 & 0 \\ \hline 0 & -1 & 0 \end{array} \right), \quad q_2 = \left(\begin{array}{cc|c} 0 & 0 & 0 \\ 0 & 0 & 1 \\ \hline 1 & 0 & 0 \end{array} \right).$$

The Lie superalgebra is characterized by a mixture of commutators and anticommutators, depending on whether an element of $osp(1|2)$ is “bosonic” (supported in the diagonal blocks of the matrix) or “fermionic” (supported in the off-diagonal blocks). In detail, the anticommutators are

$$(3.62) \quad q_1^2 = -e, \quad q_2^2 = f, \quad \{q_1, q_2\} = h,$$

and the nonzero commutators are

$$(3.63) \quad \begin{aligned} [h, e] &= 2e, & [h, f] &= -2f, & [e, f] &= h, \\ [h, q_1] &= q_1, & [h, q_2] &= -q_2, & [e, q_2] &= q_1, & [f, q_1] &= q_2. \end{aligned}$$

Symmetry groups or supergroups in Lorentz signature are always real, so we consider the real form of $\text{OSp}(1|2)$, with u, v, θ considered real. Actually, $\text{OSp}(1|2)$ contains a central element that acts as -1 on u, v, θ . This element commutes with all of the fields of BF theory (since they transform in the adjoint representation) so we do not want to regard it as a symmetry. Thus on an orientable manifold with $\mathcal{N} = 1$ supersymmetry, we should consider BF theory of the supergroup $\text{OSp}'(1|2) = \text{OSp}(1|2)/\mathbb{Z}_2$. In $\text{OSp}'(1|2)$, the group elements $\text{diag}(1, 1, -1)$ and $\text{diag}(-1, -1, 1)$ are equivalent; either one is the operator $(-1)^F$ that distinguishes bosons and fermions.

Now we want to include \mathbf{R} and \mathbf{T} and show that it is unavoidable to get $\mathbf{R}^2 = (-1)^F$ and $\mathbf{T}^2 = 1$, corresponding to a pin^- structure.⁴² As an automorphism of $\mathfrak{sl}(2, \mathbb{R})$, $\mathbf{R} : (X, Y, Z) \rightarrow (X, Y, -Z)$ corresponds to $e \leftrightarrow -f$, $h \rightarrow -h$. The extension of this to the odd generators is $q_1 \rightarrow q_2$, $q_2 \rightarrow -q_1$. So as a transformation of the Lie superalgebra, $\mathbf{R}^2 = (-1)^F$. Note that the eigenvalues of \mathbf{R} acting on q_1, q_2 are $\pm i$. It is a little tricky to analyze \mathbf{T} because it is anti-unitary (not unitary) in quantum mechanics. As a shortcut, we note that $\mathbf{T} : (X, Y, Z) \rightarrow (-X, Y, Z)$ commutes with the subgroup $\text{SO}(1, 1)$ of boosts of the YZ plane. This subgroup acts on the pair q_1, q_2 with distinct real eigenvalues and eigenvectors. So \mathbf{T} , acting on the fermionic part of the Lie superalgebra, has 1-dimensional representations, and therefore it must satisfy $\mathbf{T}^2 = 1$, not $\mathbf{T}^2 = (-1)^F$.

We go to Euclidean signature by continuing $Y \rightarrow iY$. The metric is now $ds^2 = -dX^2 + dY^2 + dZ^2$. The reflection $\mathbf{R} : (X, Y, Z) \rightarrow (X, Y, -Z)$ will still have eigenvalues $\pm i$ in acting on the fermions. In Euclidean signature, the subgroup of $\text{SO}(2, 1)$ that commutes with \mathbf{R} is the group $\text{SO}(1, 1)$ of Lorentz boosts of the $X - Z$ plane. So the eigenvectors of \mathbf{R} , acting on q_1, q_2 (or on u, v) are the eigenvectors of a nontrivial element of $\text{SO}(1, 1)$. Such an element corresponds to a hyperbolic element of $\text{SL}(2, \mathbb{R})$.

The only unorientable manifold that we will have to study in any detail is the crosscap manifold C of section 3.4.5. We recall that C has a boundary circle S and an “internal” circle S' (a one-sided geodesic), such that the monodromies U and V of a flat connection around S and S' satisfy $U = V^2$.

Suppose that acting on the triple $\begin{pmatrix} u \\ v \\ \theta \end{pmatrix}$, $U = \text{diag}(e^{\ell/2}, e^{-\ell/2}, \pm 1)$ where ℓ is a length parameter and a spin structure of \mathbf{R} or NS type corresponds

⁴²See footnote 19 in section 2.5.3 for an explanation of why \mathbf{T}^2 and \mathbf{R}^2 have opposite properties.

to a $+$ or $-$ sign.⁴³ To satisfy $V^2 = U$ where V (like R) should act with imaginary eigenvalues on q_1, q_2 , we choose the minus sign in U , and we take $V = \text{diag}(e^{\ell/2}, -e^{-\ell/2}, \pm i)$. The choice of sign corresponds to a choice of one of the two pin^- structures on C . The fact that we had to take a minus sign in U explains, in the present context, an assertion that was made in section 2.5.3: in the case of time-reversal symmetry of pin^- type, a cylinder can end on a crosscap only if its spin structure is of NS type.

The fact that V is not real means that as soon as we are on an unorientable manifold, the gravitino field of JT supergravity is not real.⁴⁴ This is a common occurrence when a Lorentz signature field theory is continued to Euclidean signature. For example, in Euclidean signature, the fermions of the Standard Model are not real.

3.5.2. Generalities about the torsion. The generalities about BF theory and the analytic and combinatorial torsion carry over directly for a semi-simple supergroup such as $\text{OSp}'(1|2)$. In particular, we anticipate that the proof in [87–89] of equivalence of the analytic and the combinatorial torsion generalizes from non-compact semi-simple groups such as $\text{SL}(2, \mathbb{R})$ to corresponding supergroups such as $\text{OSp}'(1|2)$. The proof of the topological invariance of the combinatorial torsion certainly does go through in the same way for a supergroup as for a group, as does the proof of the gluing law for the torsion. If Y is an oriented two-manifold without boundary, then the moduli space \mathcal{R} of flat connections on Y valued in a semi-simple supergroup can be given a symplectic structure that is defined by the same formula as for an ordinary group, and this determines a measure on \mathcal{R} . On such a manifold, the torsion is “trivial” in the sense that it coincides with the symplectic measure. For unorientable Y , the torsion is not equivalent to anything more elementary and we compute it using the combinatorial definition.

The definition of the combinatorial torsion in the case of a supergroup can be made by formally imitating the definitions for a bosonic group.

⁴³In $\text{SL}(2, \mathbb{R})$, we used the matrices $\pm \text{diag}(e^{\ell/2}, e^{-\ell/2})$ to describe R or NS spin structures (see eqn. (3.19)). In $\text{OSp}(1|2)/\mathbb{Z}_2$, $\text{diag}(-e^{\ell/2}, -e^{-\ell/2}, 1)$ is equivalent to $\text{diag}(e^{\ell/2}, e^{-\ell/2}, -1)$, and the latter form is sometimes convenient.

⁴⁴In terms of pin^- structures, one would explain this as follows. To build a pin^- structure on a Euclidean signature two-manifold, one starts with a rank 2 Clifford algebra $\{\gamma_\mu, \gamma_\nu\} = -2\delta_{\mu\nu}$. This algebra can be represented by 2×2 matrices, but not by real 2×2 matrices, so a fermion field coupled to the pin^- structure is not real. In Lorentz signature, one would have $\{\gamma_\mu, \gamma_\nu\} = -2\eta_{\mu\nu}$ (with $\eta = \text{diag}(-1, 1)$), and this algebra does have a representation by real 2×2 matrices, so the gravitino field in Lorentz signature can be real.

The Lie algebra $\mathfrak{osp}(1|2)$ of $\mathrm{OSP}'(1|2)/\mathbb{Z}_2$ has an invariant nondegenerate quadratic form. Just as in the bosonic world, a nondegenerate quadratic form on a vector space such as $\mathfrak{osp}(1|2)$ determines a measure on that space. The meaning of a measure on a super vector space is more subtle than in the bosonic case, but we postpone discussing that point and proceed formally for the moment. First of all, the vector spaces \mathcal{C}_q , $q = 2, 1, 0$, are defined in the same way as in section 3.2 as the direct sum of copies of the Lie algebra $\mathfrak{osp}(1|2)$, with one copy for each q -cell. Hence each \mathcal{C}_q gets a measure α_q . The boundary maps $\partial : \mathcal{C}_q \rightarrow \mathcal{C}_{q-1}$ are defined precisely as in section 3.2. The basic definition (3.14) of the torsion and the generalizations (3.15) and (3.17) that are needed when the homology is nontrivial are still valid. However, when one picks the basis vectors of a vector space, some of these basis vectors are bosonic and some are fermionic. For example, if \mathcal{C}_q has dimension $n_q|m_q$ (bosonic dimension n_q and fermionic dimension m_q), then a basis of \mathcal{C}_2 consists of n_2 bosonic vectors and m_2 fermionic ones. We might write such a basis schematically as $s_1, \dots, s_{n_2} | \tilde{s}_1, \dots, \tilde{s}_{m_2}$, where the s_i are bosonic and the \tilde{s}_j are fermionic. Similarly, to complete $\partial s_1 \cdots \partial s_{n_2} | \partial \tilde{s}_1 \cdots \partial \tilde{s}_{m_2}$ to a basis of \mathcal{C}_1 (assuming for simplicity that the homology vanishes) we add bosonic and fermionic vectors $t_1, \dots, t_{n_1-n_2} | \tilde{t}_1, \dots, \tilde{t}_{m_1-m_2}$.

All these bosonic and fermionic vectors are then included in the definition of the torsion in eqn. (3.14) and its generalizations, but some subtlety is hidden in the question of what is meant by a measure on a vector space. The basic idea is that measures for bosons transform oppositely to measures for fermions. To identify the essential point, suppose that V is a 1-dimensional bosonic vector space with measure α , and let $v \in V$ be a basis vector, and λ a complex scalar. Then $\alpha(\lambda v) = |\lambda| \alpha(v)$. But if V is a 1-dimensional fermionic vector space with measure α and basis vector ψ , then $\alpha(\lambda \psi) = \lambda^{-1} \alpha(\psi)$. More generally, suppose that V is a vector space of dimension $n|m$ with basis $s_1, \dots, s_n | \tilde{s}_1, \dots, \tilde{s}_m$. Suppose that another basis $r_1, \dots, r_n | \tilde{r}_1, \dots, \tilde{r}_m$ is obtained by acting on the first basis with a supermatrix M :

$$(3.64) \quad \begin{pmatrix} r_1 \\ \vdots \\ r_n \\ \hline \tilde{r}_1 \\ \vdots \\ \tilde{r}_m \end{pmatrix} = M \begin{pmatrix} s_1 \\ \vdots \\ s_n \\ \hline \tilde{s}_1 \\ \vdots \\ \tilde{s}_m \end{pmatrix} = \left(\begin{array}{c|c} A & B \\ \hline C & D \end{array} \right) \begin{pmatrix} s_1 \\ \vdots \\ s_n \\ \hline \tilde{s}_1 \\ \vdots \\ \tilde{s}_m \end{pmatrix},$$

where A and D are bosonic blocks and B and C are fermionic. Then the defining property of a measure α (or a density, in the terminology of [94], section 3.9; see also [95, 96]) is that it is a function on bases such that, if we denote the two bases just as \vec{r} and \vec{s} respectively, then

$$(3.65) \quad \alpha(\vec{r}) = \text{Ber}'(M)\alpha(\vec{s}).$$

This is the superspace generalization of the bosonic formula (3.13).

But we need to explain the meaning of Ber' . First of all, the superanalog of the determinant is the Berezinian, denoted Ber . Like the determinant, it obeys $\text{Ber}(M_1M_2) = \text{Ber}(M_1)\text{Ber}(M_2)$ (this is necessary for the consistency of eqn. (3.65)). There is a particularly simple formula for $\text{Ber} M$ if M is upper or lower block triangular, in the sense that one of its fermionic blocks B or C vanishes. In that case,

$$(3.66) \quad \text{Ber} M = \det A \cdot \frac{1}{\det D}.$$

In evaluating Berezinians, there is also an analog of the usual row and column reduction for determinants. This is really all we will need to know about the Berezinian.

However, if α is supposed to be a *measure* (rather than a superspace analog of a differential form of top degree), we need a slight refinement of the Berezinian. Ber' is defined for a matrix $M = \left(\begin{array}{c|c} A & B \\ \hline C & D \end{array} \right)$ that has the property that A is invertible and moreover is real modulo nilpotent variables. This being so, $\det A$ is real and nonzero modulo nilpotents and has a well-defined sign, which moreover is invariant under conjugation of M (so this sign does not depend on a specific choice of how to decompose M in bosonic and fermionic blocks). Thus we can define

$$(3.67) \quad \text{Ber}'(M) = \text{sign}(\det A) \text{Ber}(M).$$

This is the natural superanalog of $|\det M|$ in the bosonic case.⁴⁵

One difference between groups and supergroups is that the torsion for a supergroup has no obvious positivity properties, and therefore, in contrast

⁴⁵The fact that one uses an absolute value for the block A and not for its fermionic analog D is related to the following fact about Gaussian integrals. A bosonic Gaussian integral involves an absolute value; for instance the integral $\frac{1}{2\pi} \int_{-\infty}^{\infty} dx dy \exp(i\lambda xy)$ is defined for real λ , and its evaluation involves an absolute value: $\frac{1}{2\pi} \int_{-\infty}^{\infty} dx dy \exp(i\lambda xy) = 1/|\lambda|$. But an analogous fermionic integral is defined for all λ and is holomorphic in λ : $\int d\psi d\chi \exp(\lambda\psi\chi) = \lambda$.

to the bosonic case, one does need to be careful with the overall sign. This is unfortunately easier said than done. We will see that on an unorientable manifold, the torsion is actually complex-valued. This should not be too startling, because in many theories, fermionic path integrals in Euclidean signature are complex-valued. The Standard Model is an example. What is slightly unusual about super JT gravity is that its path integral is real on an orientable manifold, and becomes complex only on an unorientable manifold. This, however, is also not unique; a Majorana fermion in $2 + 1$ dimensions has the same behavior.

In section 3.5.3, we compute the torsion of an $\text{OSp}'(1|2)$ flat connection on a circle or a crosscap. These calculations are an almost immediate analog of what we have already done in the bosonic case. With these results, we can generalize the analysis of section 3.4.7 and predict the contribution in supersymmetric JT gravity of a spacetime that consists of a trumpet ending on a crosscap. To be able to compute for a general unorientable spacetime, we also need the torsion of a three-holed sphere, which is evaluated in section 3.5.4. Concluding remarks are in section 3.5.5.

3.5.3. First cases and application. A hyperbolic element U of $\text{OSp}'(1|2)$, acting on the usual triple (u, v, θ) , is of the form $\text{diag}(e^{a/2}, e^{-a/2}, \delta)$, where $\delta = 1$ or -1 for a spin structure of R or NS type and a is real. If U is the holonomy of a geodesic in a super Riemann surface with a hyperbolic metric, then the length of the geodesic is $|a|$. In computing the torsion of a flat connection on a circle S , we can assume that the holonomy is of this form.

We can proceed exactly as in section 3.4.3. S has a cell decomposition with one 0-cell and one 1-cell, so \mathcal{C}_0 and \mathcal{C}_1 are both copies of the Lie superalgebra $\text{osp}(1|2)$. As in eqn. (3.39), the boundary map is $\partial s = UsU^{-1} - s$. Moreover, as in the bosonic case, we can decompose $\text{osp}(1|2)$ as $\text{osp}(1|2)^0 \oplus \text{osp}(1|2)^\perp$, where $\text{osp}(1|2)^0$ is the Lie algebra generated by h , and $\text{osp}(1|2)^\perp$ is the orthocomplement of this. (Thus $\text{osp}(1|2)^\perp$ has dimension $2|2$, and is generated by e, f, q_1, q_2 .) The boundary operator has a similar decomposition $\partial = \partial^0 \oplus \partial^\perp$, and hence the torsion factorizes:

$$(3.68) \quad \tau_S = \tau_S^0 \cdot \tau_S^\perp.$$

Just as in the bosonic case, $\partial^0 = 0$ and hence $\tau_S^0 = da \cdot (d\varrho)^{-1}$. The interpretation of this formula is the same as it was in the bosonic case; in applications, ϱ will be a gluing parameter. Moreover, ∂^\perp is invertible, as in the bosonic case, and therefore $\tau_S^\perp = \text{Ber}'(\partial^\perp)$. This is the analog of the fact that in the bosonic case, we had $\tau_S^\perp = |\det \partial^\perp|$. The eigenvalues of ∂^\perp

acting on bosonic states are $e^a - 1$ and $e^{-a} - 1$, while its eigenvalues acting on fermions are $\delta e^{a/2} - 1$ and $\delta e^{-a/2} - 1$. So $\text{Ber}'(\partial^\perp) = |(e^a - 1)(e^{-a} - 1)|/(\delta e^{a/2} - 1)(\delta e^{-a/2} - 1) = -\delta(e^{a/4} + \delta e^{-a/4})^2$, and therefore

$$(3.69) \quad \begin{aligned} \tau_S &= -\delta(e^{a/4} + \delta e^{-a/4})^2 da (d\rho)^{-1} \\ &= \begin{cases} 4 \sinh^2 \frac{a}{4} da (d\rho)^{-1} & \text{NS spin structure} \\ -4 \cosh^2 \frac{a}{4} da (d\rho)^{-1} & \text{R spin structure.} \end{cases} \end{aligned}$$

We can treat the crosscap manifold C in the same way. For the holonomy around the boundary circle S , we take U as above but with $\delta = -1$, so $U = \text{diag}(e^{a/2}, e^{-a/2}, -1)$. For the holonomy V around the “inner” circle S' , which should satisfy $V^2 = U$, we take $V = \text{diag}(e^{a/4}, -e^{-a/4}, i)$. (This corresponds to one of the two possible pin^- structures. The second is obtained by replacing i with $-i$, which will have the effect of complex conjugating all of the following formulas.) The boundary operator ∂ is defined by the same formulas (3.45) and (3.46) as in the bosonic case. It has the familiar decomposition $\partial = \partial^0 \oplus \partial^\perp$, and the torsion factorizes again as $\tau_C = \tau_C^0 \cdot \tau_C^\perp$. The same steps as before show that $\tau_C^0 = \frac{1}{2} da \cdot (d\rho)^{-1}$ (with a factor of $1/2$ that has the same origin as before), and that $\tau_C^\perp = \text{Ber}'(\partial'_1)$, where $\partial'_1 : \text{osp}(1|2)^\perp \rightarrow \text{osp}(1|2)^\perp$ is defined by $\partial'_1(u) = VuV^{-1} - u$.

The eigenvalues of ∂'_1 on bosons are $-e^{a/2} - 1$ and $-e^{-a/2} - 1$, while its eigenvalues on fermions are $-ie^{a/4} - 1$ and $ie^{-a/4} - 1$. So $\text{Ber}'(\partial'_1) = 2 \cosh^2(\frac{a}{4}) / (1 + i \sinh \frac{a}{4}) = 2(1 - i \sinh \frac{a}{4})$ and the torsion of the crosscap (with the chosen pin^- structure) is

$$(3.70) \quad \tau_C = \left(1 - i \sinh \frac{a}{4}\right) da \cdot (d\rho)^{-1}.$$

Now we can follow the logic of section 3.4.7 and predict, in supersymmetric JT gravity, the contribution of a trumpet that ends on a crosscap. Let Z_{SJT}^T be the path integral in supersymmetric JT gravity for a trumpet T that ends on a circle of circumference a . Then the path integral measure for a double trumpet in super JT gravity is

$$(3.71) \quad Z_{\text{SJT}}^T da d\rho Z_{\text{SJT}}^T.$$

As in our discussion of ordinary JT gravity, we do not have a convenient way to incorporate the super Schwarzian mode in the framework of combinatorial torsion, so we proceed indirectly. We write \tilde{Z}_{SJT}^T for a path integral of super JT gravity on the trumpet with “super Schwarzian” boundary conditions on the outer boundary and “torsion” boundary conditions on the inner one.

Then using the gluing law for the torsion, the path integral on the double trumpet with super Schwarzian boundary conditions on both boundaries is

$$(3.72) \quad \tilde{Z}_{\text{SJT}}^T \frac{da d\rho}{4 \sinh^2 \frac{a}{4}} \tilde{Z}_{\text{SJT}}^T.$$

Comparing to (3.70) gives⁴⁶

$$(3.73) \quad \tilde{Z}_{\text{SJT}}^T = 2 \sinh \frac{a}{4} Z_{\text{SJT}}^T.$$

By the gluing law for the torsion, the supersymmetric JT path integral for a trumpet that ends on a crosscap is just $\tilde{Z}_{\text{SJT}}^T \cdot \tau_S^{-1} \tau_C$. Using the above formulas for \tilde{Z}_{SJT}^T , τ_S and τ_C , we get

$$(3.74) \quad Z_{\text{SJT}}^T \cdot \frac{1}{2} \left(-i + \frac{1}{\sinh \frac{a}{4}} \right) da$$

for the SJT path integral of a trumpet that ends on a crosscap.

As explained in section 2.6, this has to be multiplied by a factor $\exp(-i\pi N' \eta/2)$ associated with the anomaly. With $\eta = 1/2$ for a crosscap, this is $\exp(-i\pi N'/4)$. Then we have to sum over pin^- structures. The two pin^- structures give complex conjugate contributions, since flipping the pin^- structure changes the sign of η , and also replaces V in the above derivation by its complex conjugate, as a result of which all the formulas leading to eqn. (3.74) get complex conjugated.

So after summing over pin^- structures, the super JT path integral measure for a trumpet ending on a crosscap is

$$(3.75) \quad Z_{\text{SJT}}^T \cdot 2 \text{Re} \left(\exp(-i\pi N'/4) \cdot \frac{1}{2} \left(-i + \frac{1}{\sinh \frac{a}{4}} \right) \right) da \\ = Z_{\text{SJT}}^T \cdot \left(-\sin \frac{\pi N'}{4} + \frac{\cos \frac{\pi N'}{4}}{\sinh \frac{a}{4}} \right) da.$$

Here $a/2$ is the length of the one-sided geodesic on which the crosscap “ends.” The integral over a diverges for $a \rightarrow 0$, as in the bosonic theory, unless N' is congruent to 2 or 6 mod 8.

We should acknowledge a gap in this computation. Although it is true that the values of η for a pin^- structure on a crosscap are $\pm 1/2$, and that the

⁴⁶We have to take a square root here. Because of reflection positivity, we expect that we should take the positive square root.

two pin^- structures are characterized by $V = \text{diag}(e^{a/2}, -e^{-a/2}, \pm i)$, we did not carefully work out how the two choices of sign are correlated. Reversing this relationship would have the same effect as $N' \rightarrow -N'$. Actually, since the sign of the anomaly coefficient N' depends on the sign with which time-reversal acts on the elementary fermions in an underlying description such as an SYK-like model, a physically meaningful statement about the sign ultimately depends on comparing to a microscopic definition.

3.5.4. The three-holed sphere revisited. The torsion of a three-holed sphere can be computed by adapting the considerations of section 3.4.2. The holonomies U, V, W of a flat connection around the three holes satisfy $UVW = 1$. This means that we can specify U and V independently (except that we want to restrict U, V, W to be hyperbolic) and then $W = V^{-1}U^{-1}$.

The moduli space of flat $\text{OSp}'(1|2)/\mathbb{Z}_2$ connections on a three-holed sphere has dimension $3|2$. (Indeed, the pair U, V depends on $3|2 + 3|2 = 6|4$ parameters, but $3|2$ parameters are removed when we divide by conjugation, leaving $3|2$ parameters.) A convenient parametrization is to take $U = RU_0R^{-1}$, $V = RV_0R^{-1}$, where R is an arbitrary element of the gauge group and

$$(3.76) \quad \begin{aligned} U_0 &= \delta_a \left(\begin{array}{cc|c} e^{a/2} & \kappa & 0 \\ 0 & e^{-a/2} & 0 \\ \hline 0 & 0 & \delta_a \end{array} \right) \exp(\xi q_1), \\ V_0 &= \delta_b \left(\begin{array}{cc|c} e^{-b/2} & 0 & 0 \\ 1 & e^{b/2} & 0 \\ \hline 0 & 0 & \delta_b \end{array} \right) \exp(\psi q_2). \end{aligned}$$

As before, a and b are real length parameters, and $\delta_a, \delta_b = \pm 1$ represent the spin structures on the corresponding circles. The prefactors δ_a, δ_b are inessential, since we really work in $\text{OSp}'(1|2) = \text{OSp}(1|2)/\mathbb{Z}_2$, but are included to agree more smoothly with the natural formulas in $\text{SL}(2, \mathbb{R})$. The three bosonic moduli are a, b, κ , and the two fermionic moduli are ξ, ψ . We have $W = V^{-1}U^{-1} = RW_0R^{-1}$ with

$$(3.77) \quad W_0 = \delta_a \delta_b \left(\begin{array}{cc|c} e^{-(a-b)/2} & -\kappa e^{b/2} & -\delta_a \xi e^{b/2} \\ -e^{-a/2} & \kappa + e^{(a-b)/2} - \delta_b \psi \xi e^{a/2} & \delta_a \xi - \delta_a \delta_b \psi \\ \hline -\psi e^{-(a-b)/2} & \psi \kappa e^{b/2} + \delta_b \xi e^{a/2} & \delta_b \delta_a + \delta_a \psi \xi e^{b/2} \end{array} \right).$$

W will be conjugate to $\delta_c \text{diag}(e^{c/2}, e^{-c/2} | \delta_c)$ for some c and δ_c ; the length of the third hole is $|c|$, and its spin structure is controlled by $\delta_c = \pm 1$. To

determine c , we use the fact that the supertrace is invariant under conjugation.⁴⁷ So $\text{Str } W = \delta_c(2 \cosh c/2 - \delta_c)$. Setting this equal to $\text{Str } W_0$, and using the fact that $\delta_a \delta_b \delta_c = -1$, we get

$$(3.78) \quad -2 \cosh \frac{c}{2} = 2 \cosh \frac{a-b}{2} + \kappa - \psi \xi (e^{a/2} \delta_b + e^{b/2} \delta_a)$$

or

$$(3.79) \quad \kappa = -2 \cosh \frac{c}{2} - 2 \cosh \frac{a-b}{2} + \psi \xi (e^{a/2} \delta_b + e^{b/2} \delta_a).$$

The same reasoning as in the bosonic case leads to the condition

$$(3.80) \quad \text{vol}_G(U) \cdot \text{vol}_G(V) = \text{vol}_G(R) \cdot \tau,$$

which determines τ . As in the bosonic case, it suffices to impose this condition at $R = 1$.

However, we need a convenient way to describe concretely a measure such as $\text{vol}_G(U)$. Suppose that N is a super vector space of dimension $n|m$. Let

$$\vec{f} = (f^1, f^2, \dots, f^n | g^1, g^2, \dots, g^m)$$

be a basis of N , and let $(f_1, f_2, \dots, f_n | g_1, g_2, \dots, g_m)$ be the dual basis of the dual vector space N^* . Then as a tautology there is a measure on N that we denote symbolically as $\alpha = [f_1, f_2, \dots, f_n | g_1, g_2, \dots, g_m]$. It is defined by saying that the value it assigns to the basis \vec{f} is $\alpha(\vec{f}) = 1$, while the value for any other basis of N is then determined by eqn. (3.65). Under a change of basis of N^* by

$$(3.81) \quad \begin{pmatrix} f_1 \\ \vdots \\ f_n \\ \hline g_1 \\ \vdots \\ g_m \end{pmatrix} = M \begin{pmatrix} f'_1 \\ \vdots \\ f'_n \\ \hline g'_1 \\ \vdots \\ g'_m \end{pmatrix}$$

⁴⁷The supertrace of a matrix $M = \left(\begin{array}{c|c} A & B \\ \hline C & D \end{array} \right)$ is defined to be $\text{Str } M = \text{Tr } A - \text{Tr } D$.

with some matrix M , we have

$$(3.82) \quad [f_1, \dots, f_n | g_1, \dots, g_m] = \text{Ber}'(M)[f'_1, \dots, f'_n | g_1, \dots, g'_m].$$

As special cases of this, if we rescale a bosonic basis vector of the dual space, say by $f_1 \rightarrow \lambda f_1$, the measure is rescaled by a factor of $|\lambda|$,

$$(3.83) \quad \begin{aligned} & [\lambda f_1, f_2, \dots, f_n | g_1, g_2, \dots, g_m] \\ &= |\lambda| [f_1, f_2, \dots, f_n | g_1, g_2, \dots, g_m], \end{aligned}$$

but if we similarly rescale a fermionic basis vector of the dual space, say by $g_1 \rightarrow \lambda g_1$, then the measure is rescaled by a factor of λ^{-1} :

$$(3.84) \quad \begin{aligned} & [f_1, f_2, \dots, f_n | \lambda g_1, g_2, \dots, g_m] \\ &= \lambda^{-1} [f_1, f_2, \dots, f_n | g_1, g_2, \dots, g_m]. \end{aligned}$$

This λ^{-1} is the reason that, in contrast to an ordinary manifold, a measure on a supermanifold cannot be interpreted as a differential form. Eqn. (3.84) generalizes the following property of fermionic integration: if θ is a fermionic variable, then $\int d(\lambda\theta)(A + B\theta) = \lambda^{-1} \int d\theta(A + B\theta) = \lambda^{-1}B$.

A measure $\text{vol}_G(U)$ is a measure on the tangent bundle of G , whose dual space is the cotangent bundle. In section 3.4.2, to get a basis of the cotangent bundle, we expanded $U^{-1}dU$ in the basis e, f, h of \mathfrak{sl}_2 as $U^{-1}dU = (U^{-1}dU)_e e + (U^{-1}dU)_f f + (U^{-1}dU)_h h$. Then we defined the measure $\text{vol}_G(U)$ as the three-form $4(U^{-1}dU)_e (U^{-1}dU)_f (U^{-1}dU)_h$. In the present context, to get a basis of the cotangent space, we make a similar but longer expansion of $U^{-1}dU$. As it is valued in $\text{osp}(1|2)$, it can be expanded in the basis described in eqns. (3.60) and (3.61):

$$(3.85) \quad \begin{aligned} U^{-1}dU &= (U^{-1}dU)_e e + (U^{-1}dU)_f f + (U^{-1}dU)_h h \\ &+ (U^{-1}dU)_1 q_1 + (U^{-1}dU)_2 q_2. \end{aligned}$$

Then we define the measure⁴⁸

$$(3.86) \quad \text{vol}_G(U) = [(U^{-1}dU)_e, (U^{-1}dU)_f, (U^{-1}dU)_h | (U^{-1}dU)_1, (U^{-1}dU)_2].$$

The measures $\text{vol}_G(V)$ and $\text{vol}_G(R)$ are defined similarly.

At $R = 1$, we expand $R = 1 + r + \mathcal{O}(r^2)$, and expand $r = r_e e + r_f f + r_h h + r_1 q_1 + r_2 q_2$. The general form of the torsion is⁴⁹

$$(3.87) \quad \tau = F(a, b, \kappa | \xi, \psi) [da db d\kappa | d\xi d\psi]$$

with an unknown function F . The right hand side of eqn. (3.80) can be written

$$(3.88) \quad [dr_e dr_f dr_h | dr_1 dr_2] F(a, b, \kappa | \xi, \psi) [da db d\kappa | d\xi d\psi].$$

In the bosonic case, what made possible a simple computation of the torsion was that at $R = 1$, U was upper triangular and V was lower triangular; this led to simple formulas (3.33) and (3.34) for $(U^{-1}dU)_f$ and $(V^{-1}dV)_e$ as multiples of r_f and r_e . That made it possible to eliminate $(U^{-1}dU)_f$, $(V^{-1}dV)_e$, and likewise dr_e , dr_f from eqn. (3.80), and thereby to get a simple result for the torsion.

The ansatz (3.76) has been chosen to lead to similar simplifications. First, at $R = 1$, we have

$$(3.89) \quad (U^{-1}dU)_f = (e^a - 1)dr_f, \quad (V^{-1}dV)_e = (e^b - 1)dr_e.$$

These formulas enable us to eliminate from eqn. (3.80) all the 1-forms written on the left or right hand sides of eqn. (3.89). We may then set $dr_f = dr_e = 0$,

⁴⁸ This is the Riemannian measure for the Riemannian metric on the Lie algebra $\mathfrak{osp}(1|2)$ associated with the quadratic form $(x, x) = 2 \text{Str } x^2$. With $x = x_h h + x_e e + x_f f + x_1 q_1 + x_2 q_2$, we have $2 \text{Str } x^2 = 4x_h^2 + 4x_e x_f + 8x_1 x_2$. The Berezinian of the corresponding metric tensor is 1, so the Riemannian measure is $1 \cdot [dx_e dx_f dx_h | dx_1 dx_2]$. The metric tensor that we have used here restricts on $\mathfrak{sl}_2 \subset \mathfrak{osp}(1|2)$ to the metric tensor that we used to define a measure on \mathfrak{sl}_2 (see footnote 38). The latter was chosen so that the induced symplectic structure on the moduli space of Riemann surfaces is the Weil-Petersson form with standard normalization. So the torsion that we will compute using the measure (3.86) will agree with the volume form on the moduli space of super Riemann surfaces that one would get from a symplectic form on supermoduli space normalized so that on its reduced space (which is the moduli space of Riemann surfaces with spin structure) it coincides with the usual Weil-Petersson form.

⁴⁹ Here and later, when the basis vectors are 1-forms, we omit commas and write $[da db d\kappa | d\xi d\psi]$ rather than the clumsy $[da, db, d\kappa | d\xi, d\psi]$.

and compute

$$(3.90) \quad (U^{-1}dU)_2 = (\delta_a e^{a/2} - 1)dr_2, \quad (V^{-1}dV)_1 = (\delta_b e^{b/2} - 1)dr_1.$$

We can use these formulas to remove some more factors from eqn. (3.80), but in doing so, we have to take into account that rescaling dr_1 or dr_2 by a factor λ will multiply the measure by λ^{-1} , as in eqn. (3.84). After these two steps, the reduced version of eqn. (3.80) is

$$(3.91) \quad \frac{(e^a - 1)(e^b - 1)}{(\delta_a e^{a/2} - 1)(\delta_b e^{b/2} - 1)} [(U^{-1}dU)_e (U^{-1}dU)_h | (U^{-1}dU)_1] \\ \times [(V^{-1}dV)_f (V^{-1}dV)_h | (V^{-1}dV)_2] \\ = F(a, b, \kappa | \xi, \psi) [dr_h da db d\kappa | d\xi d\psi],$$

where R is set to 1 and dr_e, dr_f, dr_1 and dr_2 are set to 0. This is the analog of eqn. (3.35).

To complete the calculation, we work out the generalization of eqn. (3.36):

$$(3.92) \quad \begin{aligned} (U^{-1}dU)_h &= \frac{1}{2}da \\ (U^{-1}dU)_e &= e^{-a/2}d\kappa + \frac{\kappa}{2}e^{-a/2}da + 2\kappa e^{-a/2}dr_h + \xi d\xi \\ (U^{-1}dU)_1 &= d\xi + \frac{1}{2}da \xi + dr_h \xi \\ (V^{-1}dV)_h &= -\frac{1}{2}db \\ (V^{-1}dV)_f &= \frac{1}{2}e^{-b/2}db - 2e^{-b/2}dr_h - \psi d\psi \\ (V^{-1}dV)_2 &= d\psi + \frac{1}{2}db \psi - dr_h \psi. \end{aligned}$$

Using these results in eqn. (3.91) with the help of the change of variables formula (3.82), we finally arrive at

$$(3.93) \quad F(a, b, \kappa | \xi, \psi) = \frac{2 \sinh \frac{a}{2} \sinh \frac{b}{2}}{(\delta_a e^{a/2} - 1)(\delta_b e^{b/2} - 1)}.$$

Concretely, the steps that lead to this formula are analogous to the ones that lead to eqn. (3.37). We first replace $(U^{-1}dU)_h$ and $(V^{-1}dV)_h$ in eqn. (3.91) with $da/2$ and $-db/2$, cancel da and db in (3.91), and set $da = db = 0$ in

the remaining equations in (3.92). At this stage, eqn. (3.91) reduces to

$$(3.94) \quad \frac{\frac{1}{4}(e^a - 1)(e^b - 1)}{(\delta_a e^{a/2} - 1)(\delta_b e^{b/2} - 1)} \times [(U^{-1}dU)_e|(U^{-1}dU)_1] \cdot [(V^{-1}dV)_f|(V^{-1}dV)_2] = F(a, b, \kappa|\xi, \psi)[dr_h d\kappa|d\xi d\psi],$$

and we have $(U^{-1}dU)_1 = d\xi + dr_h \xi$ and $(V^{-1}dV)_2 = d\psi - dr_h \psi$. The Berezinian of a linear transformation from the basis of 1-forms $dr_h, d\kappa|d\xi, d\psi$ to the basis $dr_h, d\kappa|(d\xi + dr_h \xi), (d\psi - dr_h \psi)$ is 1, so on the right hand side of eqn. (3.94), we can replace $[dr_h d\kappa|d\xi d\psi]$ with $[dr_h d\kappa|(d\xi + dr_h \xi) (d\psi - dr_h \psi)]$. Then we can cancel $(U^{-1}dU)_1$ and $(V^{-1}dV)_2$ on the left of eqn. (3.94) against $(d\xi + dr_h \xi) (d\psi - dr_h \psi)$ on the right, and set $d\xi + dr_h \xi = d\psi - dr_h \psi = 0$ in the remaining equations. At this point, $(U^{-1}dU)_e(V^{-1}dV)_f$ reduces to $2e^{-(a+b)/2} d\kappa dr_h$, and eqn. (3.94) reduces to (3.93).

Instead of writing the torsion as $\tau = F(a, b, \kappa|\xi, \psi)[da db d\kappa|d\xi d\psi]$, we can use eqn. (3.78) and replace $d\kappa$ in this formula with $\sinh \frac{c}{2} dc$ (there is no sign here, because a measure as opposed to a differential form does not change sign in changing variables from κ to c). So the torsion is

$$(3.95) \quad \tau = \frac{2 \sinh \frac{a}{2} \sinh \frac{b}{2} \sinh \frac{c}{2}}{(\delta_a e^{a/2} - 1)(\delta_b e^{b/2} - 1)} [da db dc|d\xi d\psi].$$

This formula is not symmetric in a, b , and c . This happened because we defined the fermionic moduli in a way that did not treat the three boundaries symmetrically. The formula actually is invariant under a symmetry that exchanges a and b , because the ansatz that we used is invariant up to conjugation under such an operation. Interchanging $U \leftrightarrow V$ and conjugating by

$$(3.96) \quad \left(\begin{array}{cc|c} \sqrt{-\kappa} & 0 & 0 \\ 0 & \frac{1}{\sqrt{-\kappa}} & 0 \\ \hline 0 & 0 & 1 \end{array} \right) \cdot \left(\begin{array}{cc|c} 0 & -1 & 0 \\ 1 & 0 & 0 \\ \hline 0 & 0 & 1 \end{array} \right)$$

(where $-\kappa > 0$ according to eqn. (3.79)) maps $a \leftrightarrow b, \delta_a \leftrightarrow \delta_b$ and transforms the fermionic moduli by $(\xi, \psi) \rightarrow (-\psi\sqrt{-\kappa}, \xi/\sqrt{-\kappa})$, or equivalently

$$(3.97) \quad \xi \pm i\psi\sqrt{-\kappa} \rightarrow \pm i(\xi \pm i\psi\sqrt{-\kappa}).$$

The measure is invariant under this operation, as it should be since the torsion is intrinsically defined once we pick a measure on the $osp(1|2)$ Lie

algebra. It may be possible to define the fermionic moduli in a way that respects the full permutation symmetry of the three boundaries, but a convenient way to do this is not immediately obvious.

3.5.5. Assembling the pieces. Since we have computed the torsion of a flat $\text{OSp}'(1|2)/\mathbb{Z}_2$ connection on a three-holed sphere, a circle, or a cross-cap, it is straightforward now, following the same steps as in sections 3.4.4 and 3.4.6, to glue the pieces together and compute the torsion of such a connection on any two-manifold Y .

For example, we can build a closed oriented surface Y of genus g without boundary by gluing together $2g - 2$ three-holed spheres along $3g - 3$ circles. Let \mathcal{S} be the set of three-holed spheres and \mathcal{T} be the set of circles. Schematically, the gluing formula says that the torsion of Y is a simple tensor product:

$$(3.98) \quad \tau_Y = \otimes_{t \in \mathcal{T}} \tau_t \otimes_{s \in \mathcal{S}} \tau_s^{-1}$$

of the torsions of the three-holed spheres $t \in \mathcal{T}$ and the inverse torsions of circles $s \in \mathcal{S}$. The variables that the torsion depends on are a length parameter ℓ_s and a gluing parameter ρ_s for each circle $s \in \mathcal{S}$, and a pair of fermionic moduli ξ_t, ψ_t for each three-holed sphere $t \in \mathcal{T}$.

In the product (3.98), various trigonometric factors appear. According to eqn. (3.69), the inverse torsion of a circle with circumference ℓ and spin structure δ contains a factor $1/(e^{\ell/4} + \delta e^{-\ell/4})^2$. On the other hand, the s^{th} circle is a boundary of two different three-holed spheres $t, t' \in \mathcal{T}$ (exceptionally it is possible to have $t = t'$). It is natural to combine one factor $1/(e^{\ell_s/4} + \delta e^{-\ell_s/4})$ with τ_t and one with $\tau_{t'}$. The trigonometric factor that appears in the formula (3.95) for the torsion simplifies when multiplied by a product of three factors $1/(e^{\ell_s/4} + \delta_s e^{-\ell_s/4})$ in which ℓ_s is taken to be a, b , or c . The product of all those factors is

$$(3.99) \quad \frac{1}{4} \delta_a \delta_b e^{-(a+b)/4} (e^{c/4} - \delta_c e^{-c/4}).$$

Multiplying everything together, and multiplying by $\frac{1}{2}$ to take account⁵⁰ of the \mathbb{Z}_2 symmetry generated by $(-1)^F$, the result that we get for the torsion

⁵⁰If we want to think of the moduli space of flat $\text{OSp}'(1|2)$ connections as a supermanifold (with some orbifold singularities of positive bosonic codimension), then we do not want to divide the parameter space by this \mathbb{Z}_2 ; instead we take the \mathbb{Z}_2 symmetry into account by dividing τ_Y by 2.

of Y is

$$(3.100) \quad \tau_Y = \frac{1}{2}(-1)^{w_R} \prod_{s \in \mathcal{S}} [d\ell_s d\rho_s] \\ \times \prod_{t \in \mathcal{T}} \frac{1}{4} \delta_{a_t} \delta_{b_t} e^{-(a_t+b_t)/4} (e^{c_t/4} - \delta_{c_t} e^{-c_t/4}) [d\xi_t d\psi_t].$$

The parameter w_R is the number of circles $s \in \mathcal{S}$ with $\delta = 0$; this factor arises from the sign in the torsion of a circle. Eqn. (3.100) is a superanalog of the usual formula for a measure on the moduli space of ordinary Riemann surfaces, in length-twist coordinates. By gluing crosscaps onto some of the three-holed spheres, one can get an analogous formula for unorientable Y , generalizing eqn. (3.52).

In the orientable case, it is possible to obtain the same formula by symplectic methods. (In a different but somewhat similar coordinate system, the symplectic structure on the moduli space of super Riemann surfaces has been described explicitly in [97].) It is straightforward to generalize eqn. (3.100) for the case that Y is a genus g surface with n boundary components.⁵¹ The only difference is that trigonometric factors associated to the boundary circles do not cancel so neatly; for a boundary circle of length ℓ , one is left with a factor $e^{\ell/4} + \delta e^{-\ell/4}$. These factors disappear if the external boundary circles have NS spin structure ($\delta = -1$) and we glue them onto trumpets, In that case, according to eqn. (3.73), we get a factor $2 \sinh \frac{\ell}{4}$ in converting from a “torsion” path integral \tilde{Z}_{SJT} on the trumpet to a conventional SJT path integral Z_{SJT} ; gluing via τ_S^{-1} gives a factor $1/4 \sinh^2 \frac{\ell}{4}$, according to eqn. (3.69). The product of these factors with $e^{\ell/4} + \delta e^{-\ell/4}$ is simply 1.

At first, the formula (3.15) might make one suspect that for any surface Y obtained by gluing of three-holed spheres (possibly with crosscaps), the moduli space of flat connections on Y would have zero volume, on the grounds that the volume form in eqn. (3.15) does not depend on the odd moduli, and so appears to vanish upon integration over those variables. The fallacy in this argument is that the region of length-twist coordinates that describes the moduli space of super Riemann surfaces is defined by inequalities that involve the fermionic coordinates as well as the bosonic ones. To see why this can lead to a nonzero integral, consider a toy problem with one bosonic variable κ , two fermionic variables ξ, ψ , and a measure $d\kappa d\xi d\psi$ (or $[d\kappa|d\xi d\psi]$, to be more pedantic). The volume of the region $0 \leq \kappa \leq 1$ is 0,

⁵¹When Y has a boundary, the torsion is not equivalent to the symplectic volume, since the identity (3.6) does not hold.

as one sees by integrating first over κ :

$$(3.101) \quad \int_{0 \leq \kappa \leq 1} d\kappa d\xi d\psi \cdot 1 = \int d\xi d\psi \cdot 1 = 0.$$

But suppose the integration region is $0 \leq \kappa \leq 1 + \xi\psi$. Then integrating first over κ , we get

$$(3.102) \quad \int_{0 \leq \kappa \leq 1 + \xi\psi} d\kappa d\xi d\psi \cdot 1 = \int d\xi d\psi (1 + \xi\psi) = 1.$$

In fact, we will encounter a phenomenon somewhat like that in generalizing Mirzakhani’s results to super Riemann surfaces.

The volumes do, however, vanish in genus 0, that is if Y is a hyperbolic sphere with n holes. This is explained in appendix A. In the context of JT supergravity, this vanishing implies that a connected correlator $\langle (\text{Tr} \exp(-\beta H))^n \rangle_c$ vanishes in genus 0, and receives contributions only in higher orders in the topological expansion. The analogous statement is certainly not true for ordinary JT gravity.

4. Loop equations

In this section, we will explain a technique called the “loop equations” for analyzing the $1/L$ expansion of matrix integrals. This technology will make it possible to compare matrix integrals to JT gravity and supergravity in section 5 below.

The loop equations have been studied since the work of Migdal in [47], but a significant streamlining was achieved in the work of Eynard for the case of $\beta = 2$ ensembles [39]. Using a dispersion relation method, it was shown how to turn the loop equations into a recursion relation that closes on correlation functions of resolvents. A nice feature of the resulting equations is that they do not depend explicitly on the potential of the matrix integral, only on the “spectral curve” or leading density of eigenvalues. The recursive step in this formalism is a contour integral that reduces to a sum of residues: the particular structure that appears was later abstracted as the formalism of “topological recursion” [38].

For the case of the general β ensembles, some of the simplifications of [39] are still possible. But the resulting recursion relation is somewhat more complicated: the recursive step involves an integral around a cut, rather than a sum of residues. This has been studied from a somewhat different perspective in works including [54, 55]. In section 4.1, we will explain the

loop equations for general β , pointing out the simplification that occurs for $\beta = 2$. We will limit ourselves to so-called “one-cut” models, where the large L density of eigenvalues is supported in a single interval of the real axis.

For the Altland-Zirnbauer (α, β) ensembles, the basic method of [39] is again possible. For a special case $(\alpha, \beta) = (1, 2)$, the recursion reduces to a residue computation that is equivalent to a form of “topological recursion.” For more general values of (α, β) , the recursive step involves a nontrivial integral. In section 4.2, we will explain the loop equations for these ensembles.

4.1. Loop equations for the β ensembles

As described in section 2.1.1, the Dyson β ensembles are characterized by a measure for eigenvalues of the form

$$(4.1) \quad \prod_{1 \leq i < j \leq L} |\lambda_i - \lambda_j|^\beta \prod_{i=1}^L e^{-L \frac{\beta}{2} V(\lambda_i)} d\lambda_i.$$

where we included a general potential function $V(\lambda)$ in addition to the measure factor described in section 2.1.2. $V(x)$ and β should be thought of as parameters of the distribution. As a probability measure, it makes sense for continuous values of β , but $\beta = 1, 2, 4$, are the values relevant for random matrix theory. The potential V should be a function that grows rapidly enough at infinity that the integral converges. The manipulations below will assume that V is analytic in a neighborhood of the region where the classical (large L) eigenvalue distribution is nonzero.

The goal is to set up machinery for computing the $1/L$ expansion of expectation values in the ensemble defined by (4.1). As a basis for “single-trace” observables, we will use the resolvent. This is a function that depends on a parameter x in addition to all of the eigenvalues $\{\lambda\}$, and it is defined by

$$(4.2) \quad R(x, \{\lambda\}) = \sum_{i=1}^L \frac{1}{x - \lambda_i}.$$

In what follows, we will leave the λ dependence of the resolvent implicit, and write simply $R(x)$. From this quantity, any other single trace quantity $\sum_{i=1}^L f(\lambda_i)$ can be obtained. For example, one can start by computing the

density of eigenvalues from the discontinuity of $R(x)$ across the real axis:

$$(4.3) \quad \rho(x) = \sum_{i=1}^L \delta(x - \lambda_i) = -\frac{1}{2\pi i} (R(x + i\epsilon) - R(x - i\epsilon)).$$

Then one can write $\sum_{i=1}^L f(\lambda_i) = \int dx \rho(x) f(x)$.

Below, we will use a shorthand notation for products of resolvents

$$(4.4) \quad R(x_1, \dots, x_n) = R(x_1)R(x_2) \dots R(x_n)$$

and further abbreviate by writing the LHS as $R(I)$, where

$$(4.5) \quad I = \{x_1, \dots, x_n\}, \quad |I| = n.$$

Connected correlation functions of resolvents have an asymptotic expansion in powers of $1/L$, which we can write using the above notation:

$$(4.6) \quad \langle R(I) \rangle_c \simeq \sum_{g=0, \frac{1}{2}, 1, \dots} \frac{R_g(I)}{L^{2g+|I|-2}}.$$

This expression defines the quantities $R_g(I)$ as the coefficients in the $1/L$ expansion. In general, the “genus” summation index g takes both integer and half-integer values, although we will see that only integer values contribute for the special case $\beta = 2$.

The loop equations are simply the observation that

$$(4.7) \quad 0 = \int_{-\infty}^{\infty} d^L \lambda \frac{\partial}{\partial \lambda_a} \left[\frac{1}{x - \lambda_a} R(I) \prod_{i < j} |\lambda_i - \lambda_j|^\beta \prod_i e^{-L \frac{\beta}{2} V(\lambda_i)} \right],$$

where $a \in \{1, \dots, L\}$ is some fixed index. Distributing the derivative, we find

$$(4.8) \quad 0 = \left\langle \left[\frac{1}{(x - \lambda_a)^2} + \beta \frac{1}{x - \lambda_a} \sum_{j \neq a} \frac{1}{\lambda_a - \lambda_j} - \frac{L\beta}{2} \frac{V'(\lambda_a)}{x - \lambda_a} \right] R(I) + \frac{1}{x - \lambda_a} \partial_{\lambda_a} R(I) \right\rangle.$$

Here the angle brackets mean an expectation value in the ensemble defined by (4.1). Because the eigenvalue distribution is permutation-invariant, we do not learn any new information by setting a to any particular value, and

below we will sum over a . As much as possible, we would like to reduce (4.8) to an expression involving expectation values of products of resolvents. We will do this in three steps.

(i) The second term inside the brackets in (4.8) can be rewritten using

$$(4.9) \quad \frac{1}{x - \lambda_a} \frac{1}{\lambda_a - \lambda_j} \rightarrow \frac{1}{2} \left(\frac{1}{x - \lambda_a} \frac{1}{\lambda_a - \lambda_j} + \frac{1}{x - \lambda_j} \frac{1}{\lambda_j - \lambda_a} \right)$$

$$(4.10) \quad = \frac{1}{2} \frac{1}{x - \lambda_a} \frac{1}{x - \lambda_j}.$$

In the first step, we replaced the LHS by its symmetrization under $a \leftrightarrow j$. Summing over a , we can then rewrite the first two terms inside brackets in (4.8) as

$$(4.11) \quad \sum_a \frac{1}{(x - \lambda_a)^2} + \frac{\beta}{2} \sum_{a,j \neq a} \frac{1}{x - \lambda_a} \frac{1}{x - \lambda_j}$$

$$= \sum_a \frac{1 - \frac{\beta}{2}}{(x - \lambda_a)^2} + \frac{\beta}{2} \sum_{a,j} \frac{1}{x - \lambda_a} \frac{1}{x - \lambda_j}$$

$$(4.12) \quad = -(1 - \frac{\beta}{2}) \partial_x R(x) + \frac{\beta}{2} R(x, x).$$

(ii) The term involving the potential cannot be completely rewritten in terms of resolvents, but we can improve the situation by writing

$$(4.13) \quad \sum_a \frac{V'(\lambda_a)}{x - \lambda_a} R(I) = V'(x) R(x, I) - P(x; I)$$

where we define

$$(4.14) \quad P(x; I) \equiv \sum_a \frac{V'(x) - V'(\lambda_a)}{x - \lambda_a} R(I).$$

The first term on the LHS of (4.13) is now written in terms of resolvents. The remainder term $P(x; I)$ is not, but a key point is that $P(x; I)$ is analytic in x (assuming V is itself analytic). For example, if V is a polynomial, then P is also a polynomial. Below, we will use a dispersion relation that will allow us to ignore the analytic $P(x; I)$ term.

(iii) Finally, the last term in (4.8) can be rewritten as follows. We start with

$$(4.15) \quad \partial_{\lambda_a} R(I) = \sum_{k=1}^n \frac{1}{(x_k - \lambda_a)^2} R(I \setminus x_k)$$

where $I \setminus x_k$ means the coordinates $\{x_1, \dots, x_n\}$ with x_k removed. Then using

$$(4.16) \quad \frac{1}{(x - \lambda_a)(x_k - \lambda_a)^2} = \partial_{x_k} \frac{\frac{1}{x - \lambda_a} - \frac{1}{x_k - \lambda_a}}{x - x_k}$$

and summing over a , we have

$$(4.17) \quad \sum_{a=1}^L \frac{1}{x - \lambda_a} \partial_{\lambda_a} R(I) = \sum_{k=1}^n \partial_{x_k} \frac{R(x, I \setminus x_k) - R(I)}{x - x_k}.$$

Putting these steps together, and multiplying through by $\frac{2}{\beta}$, one can rewrite (4.8) as

$$(4.18) \quad \left\langle \begin{aligned} & \left(1 - \frac{2}{\beta}\right) \partial_x R(x, I) + R(x, x, I) \\ & - LV'(x)R(x, I) + \frac{2}{\beta} \sum_{k=1}^n \partial_{x_k} \frac{R(x, I \setminus x_k) - R(I)}{x - x_k} \end{aligned} \right\rangle = -L \langle P(x; I) \rangle.$$

Next, we would like to write (4.18) in terms of connected correlation functions, and use the lower-order equations to simplify. This is a little tricky. It is helpful to focus on a proper subset $J \subsetneq I$ and collect all terms that multiply $\langle R(J) \rangle_c$. One can show that the sum of all terms that multiply this expression is precisely the quantity that vanishes due to (4.18) with $I \rightarrow I \setminus J$. So in writing (4.18) in terms of connected correlators, we can omit any terms that have a factor of $\langle R(J) \rangle_c$ with $J \subsetneq I$. The same argument allows us to omit the terms involving $\langle R(I) \rangle_c$ except for the one that arises from the last term on the first line of (4.18). Writing out what remains, we have

$$(4.19) \quad \begin{aligned} & \left(1 - \frac{2}{\beta}\right) \partial_x \langle R(x, I) \rangle_c + \langle R(x, x, I) \rangle_c \\ & + \sum_{J \subsetneq I} \langle R(x, J) \rangle_c \langle R(x, I \setminus J) \rangle_c - LV'(x) \langle R(x, I) \rangle_c \\ & + \frac{2}{\beta} \sum_{k=1}^n \partial_{x_k} \left(\frac{\langle R(x, I \setminus x_k) \rangle_c - \langle R(I) \rangle_c}{x - x_k} \right) = -L \langle P(x; I) \rangle_c. \end{aligned}$$

We will now discuss the application of this equation to three special cases, before treating the generic case below.

4.1.1. A first special case $R_0(x)$, and the spectral curve. First, we consider $I = \emptyset$, and further simplify the equation by inserting (4.6) and keeping only the leading terms, proportional to L^2 . Then (4.19) reduces to

$$(4.20) \quad R_0(x)^2 - V'(x)R_0(x) = (\text{analytic in } x)$$

where the analytic RHS comes from the $P(x)$ term. After adding an analytic quantity $V'(x)^2/4$ to both sides, we can write this as

$$(4.21) \quad y^2(x) = (\text{analytic in } x)$$

where y is defined by

$$(4.22) \quad R_0(x) = \frac{V'(x)}{2} + y(x).$$

Eqn. (4.21) defines a hyperelliptic curve, with two sheets differing by the sign of y . This is referred to as the spectral curve of the matrix integral.

The quantity y has a physical interpretation as the leading approximation to the density of eigenvalues. To see this, recall that in general, the resolvent is a multivalued function, with a discontinuity across the real axis given by the density of eigenvalues (4.3). So, in particular, the discontinuity of the genus-zero resolvent $R_0(x)$ gives the leading approximation to the density of eigenvalues:

$$(4.23) \quad R_0(x + i\epsilon) - R_0(x - i\epsilon) = -2\pi i \rho_0(x), \quad \rho_0(x) = \lim_{L \rightarrow \infty} \frac{1}{L} \langle \rho(x) \rangle.$$

Here, $\rho_0(x)$ is normalized so that $\int dx \rho_0(x) = 1$. We will consider the simplest “one-cut” matrix integrals, for which $\rho_0(x)$ is supported in a single interval of the real axis $a_- \leq x \leq a_+$. Resolvents are naturally defined with a branch cut that coincides with this interval. Since $R_0(x)$ and $y(x)$ differ only by an analytic term $\frac{1}{2}V'(x)$, the same is true for $y(x)$, and we can rewrite the LHS of (4.23) as the discontinuity of $y(x)$ across the real axis. Because y^2 is analytic, the only possible discontinuity in y is a change of sign. This implies that if we approach the cut from one side or the other, $y(x)$ has to be one-half of the value of its discontinuity:

$$(4.24) \quad y(x \pm i\epsilon) = \mp i\pi \rho_0(x).$$

So, along the real axis and inside the region of support of $\rho_0(x)$, $y(x)$ is pure imaginary, with a sign that depends on which half-plane we approach from.

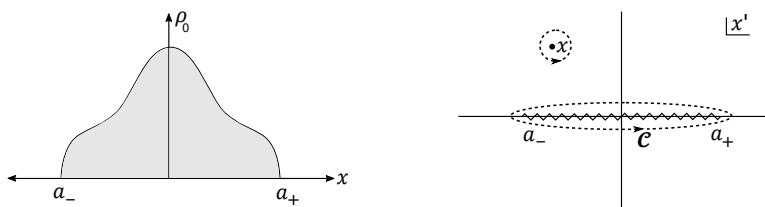


Figure 15: At left we show a typical $\rho_0(x)$ function, supported between endpoints $x = a_{\pm}$. At right we show the x' plane for the manipulation in (4.40). The original contour surrounds the point x , and the final contour surrounds the cut that runs between $x' = a_{\pm}$. For the special case of an ensemble with $\beta = 2$, the cut is replaced by poles at the endpoints.

Outside the region of support, $y(x)$ is real. At each endpoint of the cut, $y(x)$ has a square-root singularity $\sqrt{a_{\pm} - x}$.

In principle, one can solve for $y(x)$ starting from the potential $V(x)$ that defines the matrix integral. But, as we will see below, in the recursion relation we will not need to know $V(x)$ directly, only y . So in practice it is more convenient to think about things the other way around: we specify the matrix integral by giving the function $y(x)$ rather than the potential.

The resolvent can be expressed in terms of the eigenvalue density:

$$(4.25) \quad R_0(x) = \int_{a_-}^{a_+} d\lambda \frac{\rho_0(\lambda)}{x - \lambda}.$$

This integral is manifestly holomorphic in x on the complement of the interval $[a_-, a_+]$, where it has a cut. To be more exact, the integral defines $R_0(x)$ on what is known as the “first sheet.” The formula $R_0(x) = V'(x)/2 + y(x)$, where $y(x)$ is defined on the spectral curve, a double cover of the x plane, shows that $R_0(x)$ can be continued through the cut onto a second sheet. Multiresolvents can similarly be expressed in terms of a joint eigenvalue density $\rho(\lambda_1, \lambda_2, \dots, \lambda_k)$. In lowest order in $1/L$, the joint density is just a product $\rho_0(\lambda_1)\rho_0(\lambda_2)\cdots\rho_0(\lambda_k)$, but in higher orders there are corrections. Thus in general

$$(4.26) \quad R_g(I) = \int_{a_-}^{a_+} \frac{d\lambda_1}{x_1 - \lambda_1} \cdots \frac{d\lambda_n}{x_n - \lambda_n} \rho_g(\lambda_1, \lambda_2, \dots, \lambda_n),$$

where $\rho_g(\lambda_1, \lambda_2, \dots, \lambda_N)$ is the genus g contribution to $\rho(\lambda_1, \lambda_2, \dots, \lambda_n)$. Eqn. (4.26) defines a function that is manifestly holomorphic in x_1, x_2, \dots, x_n as long as they are on the complement of the interval $[a_-, a_+]$. Moreover, the

formula determines the behavior of this function when any of the x_i become large. For any reasonably behaved function $\rho_g(\lambda_1, \lambda_2, \dots, \lambda_n)$, eqn. (4.26) implies that $R_g(I)$ vanishes at least as $1/x_i$ for $x_i \rightarrow \infty$. But since the total number of eigenvalues is a fixed constant L , with no fluctuations or higher order corrections, it actually follows that $R_g(I)$ vanishes as $1/x_i^2$ or faster, except for the special case $g = 0$ and $|I| = 1$. In what follows, in writing dispersion relations for the multiresolvents, we begin on the “first sheet,” on which $R_g(I)$ has the properties just stated. However, from the analysis, we will learn that like $R_0(x)$, the functions $R_g(I)$ can be continued (in each variable separately) to a second sheet, on which in general there are additional singularities. For example, on the second sheet, $R_0(x_1, x_2)$ has a double pole at $x_1 = x_2$.

The reader might ask how we know that the functions $\rho_g(\lambda_1, \lambda_2, \dots, \lambda_n)$ have the same support as the product $\rho_0(\lambda_1)\rho_0(\lambda_2)\cdots\rho_0(\lambda_n)$. One answer comes from the fact that the procedure of topological recursion that we will describe leads to a unique answer with this property. We can also reason as follows. The fact that the eigenvalues are supported in some interval is stable against small changes in the effective potential V . But might $1/L^2$ corrections change the precise interval $[a_-, a_+]$? In a sense, this does happen, and it is reflected in the fact that $R_g(I)$ as computed by topological recursion has increasingly strong singularities at the endpoints of the interval as g or $|I|$ is increased. The Taylor series expansion $\sqrt{\lambda - a_- + c/L^2} = \sqrt{\lambda - a_-} + \frac{c}{2L^2}(\lambda - a_-)^{-1/2} + \mathcal{O}(1/L^4)$ illustrates the point: a perturbative shift in the endpoint of the cut is equivalent, in the $1/L$ expansion, to keeping the endpoints of the cut fixed and generating increasingly severe singularities at the endpoints.

4.1.2. A second special case $R_0(x, x_1)$. Next, we consider the case $I = \{x_1\}$ and again keep only the leading terms in the $1/L$ expansion, this time proportional to L . Then (4.19) gives

$$(4.27) \quad \begin{aligned} & 2R_0(x)R_0(x, x_1) - V'(x)R_0(x, x_1) \\ & + \frac{2}{\beta} \partial_{x_1} \frac{R_0(x) - R_0(x_1)}{x - x_1} = (\text{analytic in } x). \end{aligned}$$

Using (4.22), and moving some analytic terms to the RHS, this can be rewritten as

$$(4.28) \quad \begin{aligned} & 2y(x)R_0(x, x_1) + \frac{2}{\beta} \frac{y(x)}{(x - x_1)^2} \\ & = (\text{analytic in } x \text{ in neighborhood of cut}). \end{aligned}$$

Here we are assuming that x_1 is away from the cut, so we allowed ourselves to move terms to the RHS that are singular at $x = x_1$. We would like to use this equation to determine $R_0(x, x_1)$ using a dispersion relation. In general, $y(x)$ could be growing rapidly at infinity, so before we do this, one would like to simplify the equation by dividing by $y(x)$. However, since $1/y$ is singular at the endpoints of the cut, this will make some of the terms on the RHS singular. Instead, we proceed as follows. We write

$$(4.29) \quad \sigma(x) = (x - a_+)(x - a_-)$$

and define $\sqrt{\sigma(x)}$ with a branch cut running between the endpoints of the cut a_{\pm} . Then $\sqrt{\sigma(x)}/y(x)$ is analytic in a neighborhood of the cut, and we can safely multiply (4.28) by this quantity, to obtain

$$(4.30) \quad R_0(x, x_1)\sqrt{\sigma(x)} + \frac{1}{\beta} \frac{\sqrt{\sigma(x)}}{(x - x_1)^2} = (\text{analytic in } x \text{ in neighborhood of cut}).$$

Now, to determine $R_0(x, x_1)$, we write a dispersion relation

$$(4.31) \quad \begin{aligned} R_0(x, x_1)\sqrt{\sigma(x)} &= \frac{1}{2\pi i} \oint_x \frac{dx'}{x' - x} R_0(x', x_1)\sqrt{\sigma(x')} \\ &= -\frac{1}{2\pi i} \int_{\mathcal{C}} \frac{dx'}{x' - x} R_0(x', x_1)\sqrt{\sigma(x')} \\ &= \frac{1}{2\pi i} \int_{\mathcal{C}} \frac{dx'}{x' - x} \frac{1}{\beta} \frac{\sqrt{\sigma(x')}}{(x' - x_1)^2}. \end{aligned}$$

In the first equality, we used the Cauchy residue formula. In the second equality, we deformed the contour to surround the cut, see figure 15. To justify this deformation, we use our knowledge that $R_0(x, x_1)$ is holomorphic in x away from the cut. In the final step we used (4.30), taking advantage of the fact that terms that are analytic in a neighborhood of the cut will not contribute. After evaluating the integral, one finds

$$(4.32) \quad R_0(x_1, x_2) = \frac{1}{\beta(x_1 - x_2)^2} \left(\frac{x_1 x_2 - \frac{a_+ + a_-}{2}(x_1 + x_2) + a_+ a_-}{\sqrt{\sigma(x_1)}\sqrt{\sigma(x_2)}} - 1 \right).$$

An important feature of this expression is that it is regular at $x_1 = x_2$ if both coordinates are on the same sheet, but has a double pole if they are

on opposite sheets:

$$\begin{aligned}
 (4.33) \quad R_0(x, x) &= \frac{(a_+ - a_-)^2}{8\beta(a_+ - x)^2(a_- - x)^2}, \\
 R_0(\hat{x}_1, x'_2) &= -\frac{2}{\beta(x_1 - x_2)^2} + (\text{regular}).
 \end{aligned}$$

Here we are using the notation \hat{x} to mean that we continue the x coordinate to position x on the second sheet.

It will be useful below to know eqn. (4.32) in a limit in which $a_+ \rightarrow \infty$ with $a_0 = 0$. This is

$$(4.34) \quad R_0(x_1, x_2) = \frac{1}{2\beta\sqrt{-x_1}\sqrt{-x_2}(\sqrt{-x_1} + \sqrt{-x_2})^2}.$$

4.1.3. A third special case $R_0(x, x_1, x_2)$. As a final special case, we take $I = \{x_1, x_2\}$ and again work at leading order in $1/L$, which is now L^0 . Then (4.19) gives

$$\begin{aligned}
 (4.35) \quad &2y(x)R_0(x, x_1, x_2) + 2R_0(x, x_1)R_0(x, x_2) \\
 &+ \frac{2}{\beta} \left(\frac{R_0(x, x_2)}{(x - x_1)^2} + \frac{R_0(x, x_1)}{(x - x_2)^2} \right) \sim 0.
 \end{aligned}$$

where the \sim means up to terms that are analytic in x in a neighborhood of the cut. After multiplying through by $\sqrt{\sigma(x)}/(2y(x))$, one can write $R_0(x, x_1, x_2)$ using a dispersion relation as in (4.31):

$$\begin{aligned}
 (4.36) \quad R_0(x, x_1, x_2)\sqrt{\sigma(x)} &= \frac{1}{2\pi i} \int_{\mathcal{C}} \frac{dx'}{x' - x} \frac{\sqrt{\sigma(x')}}{y(x')} \\
 &\times \left(R_0(x', x_1)R_0(x', x_2) + \frac{1}{\beta} \frac{R_0(x', x_2)}{(x' - x_1)^2} + \frac{1}{\beta} \frac{R_0(x', x_1)}{(x' - x_2)^2} \right).
 \end{aligned}$$

One can check from the explicit formula (4.32) that the integrand is actually meromorphic in a neighborhood of the cut, having a simple pole at each endpoint. So the integral reduces to a sum of two residues.

4.1.4. The generic case. Now we consider the generic case. To do so, we plug the genus expansion (4.6) into the loop equations (4.19), and collect the terms that appear at order L^{2-2g-n} , where n is the number of elements

in the set $I = \{x_1, \dots, x_n\}$. The resulting equation can be written as

$$(4.37) \quad 2y(x)R_g(x, I) + F_g(x, I) \sim 0$$

where

$$(4.38) \quad F_g(x, I) = (1 - \frac{2}{\beta})\partial_x R_{g-\frac{1}{2}}(x, I) + R_{g-1}(x, x, I) + \sum_{\text{stable}} R_h(x, J)R_{g-h}(x, I \setminus J)$$

$$(4.39) \quad + 2 \sum_{k=1}^n \left(R_0(x, x_k) + \frac{1}{\beta} \frac{1}{(x-x_k)^2} \right) R_g(x, I \setminus x_k).$$

As before, the \sim symbol in (4.37) means equality up to terms that are analytic in x in a neighborhood of the cut. Working modulo such terms allowed us to drop the expressions involving $P(x; I)$ and $\frac{1}{(x-x_k)^2}R(I)$ in (4.19).

Naively, the sum on line (4.38) should be over integer and half-integer h satisfying $0 \leq h \leq g$ and over subsets $J \subseteq I$. However, the subscript “stable” means that we omit the special cases where one of the factors is $R_0(x)$ or $R_0(x, x_k)$. These special cases are treated separately: the contribution where one factor is $R_0(x)$ combines with the $-LV'(x)$ term on the second line of (4.19) to form the combination $2y(x)R_g(x, I)$ present in (4.37). The contribution where one of the factors is $R_0(x, x_k)$ combines with the $R(x, I \setminus x_k)$ term in (4.19) to give the final term on the second line of (4.39).

Eq. (4.37) determines the parts of $y(x)R_g(x, I)$ that are singular at the cut, and knowledge of these terms allows us to recover the full $R_g(x, I)$ using a dispersion relation as in (4.31):

$$(4.40) \quad R_g(x, I)\sqrt{\sigma(x)} = \frac{1}{2\pi i} \oint_x \frac{dx'}{x' - x} R_g(x', I)\sqrt{\sigma(x')}$$

$$= -\frac{1}{2\pi i} \int_{\mathcal{C}} \frac{dx'}{x' - x} R_g(x', I)\sqrt{\sigma(x')}$$

$$(4.41) \quad = \frac{1}{2\pi i} \int_{\mathcal{C}} \frac{dx'}{x' - x} F_g(x', I) \frac{\sqrt{\sigma(x')}}{2y(x')}.$$

In the second equality, we deformed the contour as in figure 15. The properties needed to justify this contour deformation were argued at the end of section 4.1.1: $R_g(x, I)$ is holomorphic in x away from the cut and vanishes as $1/x^2$ or faster at $x = \infty$. It can further be proved inductively using (4.41) that on the first sheet, $R_g(x, I)$ is actually holomorphic in x except at the

endpoints of the cut (and hence can be continued through the cut). This is equivalent to saying that the eigenvalue distribution $\rho_g(I)$ in eqn. (4.26) is real analytic except at the endpoints.

The relation (4.40), together with the special cases discussed above, gives a recursion that determines all $R_g(x, I)$.

4.1.5. Why $\beta = 2$ is special. When $\beta = 2$, the first term in (4.39) is absent. This leads to two simplifications. First, it is easy to see that this implies that all $R_g(I)$ with half-integer g vanish, so the $1/L$ expansion becomes a $1/L^2$ expansion. Second, the contour integral around the cut in (4.41) reduces to a sum of residues at the endpoints.

The second point is due to the fact that for $\beta = 2$, the multivaluedness of the $R_g(I)$ becomes very simple. We will use the notation \hat{x} in the argument of a function to indicate a coordinate that has been continued to location x but on the second sheet. So, for example, $y(\hat{x}) = -y(x)$ and $\sqrt{\sigma(\hat{x})} = -\sqrt{\sigma(x)}$. Then with the exception of $R_0(x_1)$ and $R_0(x_1, x_2)$, we claim that

$$(4.42) \quad R_g(\hat{x}_1, x_2, \dots, x_n) = -R_g(x_1, x_2, \dots, x_n).$$

This can be proven by induction. First we note that if (4.42) holds, then $F_g(x, I)$ is actually single-valued in x . This is obvious for the terms $R_{g-1}(x, x, I)$ and $R_h(x, J)R_{g-h}(x, I \setminus J)$ that appear in (4.39). It would not be true for the $\partial_x R_{g-\frac{1}{2}}(x, I)$ term, but this is absent for $\beta = 2$. For the final term in (4.39), it is true using the explicit formula (4.32).

Next, we note that single-valuedness of F_g implies that the integrand in (4.41) is itself single-valued, because $\sqrt{\sigma(x)}/y(x)$ is also single-valued. So, assuming (4.42), one finds that the contour integral around the cut just picks up the contribution of poles at the endpoints of the cut. The result of this integral will be analytic in x , and therefore $R_g(x, I)\sqrt{\sigma(x)}$ is analytic in x , so $R_g(x, I)$ satisfies (4.42). To complete the induction, we also need to check explicitly the special base case $R_0(x, x_1, x_2)$, which amounts to showing that the sum of the second and third terms from (4.35) are single-valued in x . This follows from (4.32).

The fact that the integral (4.41) reduces to residues at the endpoints of the cut is a major simplification. In practical terms, it makes the loop equations much easier to implement. Conceptually, it makes the loop equations for $\beta = 2$ a special case of the Eynard-Orantin “topological recursion” [38].

4.1.6. Relation between $\beta = 1$ and $\beta = 4$. The $\beta = 1$ and $\beta = 4$ ensembles are very closely related to each other [60]. The relationship is clearest

if, on the $\beta = 4$ side, we work in terms of \tilde{R} defined by (assuming L even)

$$(4.43) \quad \tilde{R}(x; L) = 2R(x; \frac{L}{2}), \quad \tilde{R}(I; L) = 2^{|I|}R(I; \frac{L}{2}).$$

Here we temporarily made the L argument explicit. One can interpret $\tilde{R}(x)$ as $\text{Tr} \frac{1}{x-H}$ where H is a matrix that has $\frac{L}{2}$ independent eigenvalues from the $\beta = 4$ ensemble, each with two-fold degeneracy. This is actually the situation that describes a ensemble of matrices with $\text{Sp}(L)$ symmetry: such matrices have $\beta = 4$ statistics and twofold degeneracy. We can write a $1/L$ expansion of connected correlators of $\tilde{R}(I)$ as

$$(4.44) \quad \langle \tilde{R}(I; L) \rangle_c \simeq \sum_g \frac{\tilde{R}_g(I)}{L^{2g+|I|-2}} = 2^{|I|} \sum_g \frac{R_g(I)}{(\frac{L}{2})^{2g+|I|-2}}$$

which implies that

$$(4.45) \quad \tilde{R}_g(I) = 2^{2(g+|I|-1)}R_g(I).$$

Substituting (4.45) into (4.38), one finds that the \tilde{R}_g quantities for the $\beta = 4$ ensemble satisfy

$$(4.46) \quad 2y(x)\tilde{R}_g(x, I) + \tilde{F}_g(x, I) \sim 0$$

where

$$(4.47) \quad \tilde{F}_g(x, I) = \partial_x \tilde{R}_{g-\frac{1}{2}}(x, I) + \tilde{R}_{g-1}(x, x, I) + \sum_{\text{stable}} \tilde{R}_h(x, J)\tilde{R}_{g-h}(x, I \setminus J)$$

$$(4.48) \quad + 2 \sum_{k=1}^n \left(\tilde{R}_0(x, x_k) + \frac{1}{(x-x_k)^2} \right) \tilde{R}_g(x, I \setminus x_k).$$

Now, the point is that this is the same recursion satisfied by R_g for the $\beta = 1$ ensemble, except for the sign of the first $\partial_x R_{g-\frac{1}{2}}$ term (which would have been -1 for the $\beta = 1$ ensemble). In the solution to the recursion, changing this sign has the effect of reversing the sign of all terms with half-integer g . In terms of 't Hooft double-line diagrams (or in the JT gravity interpretation below), these correspond to unorientable surfaces with an odd number of crosscaps, so we conclude that the difference between the $\beta = 1$ and $\beta = 4$ ensembles is to insert a relative factor of $(-1)^{n_c}$ where n_c is the number of crosscaps [60]. Related matters are discussed in appendix B.

4.2. Loop equations for the (α, β) ensembles

In this section, we will study the loop equations for the (α, β) ensembles of Altland and Zirnbauer [44].⁵² As discussed in section 2.1.2, these ensembles are characterized by a measure

$$(4.49) \quad \prod_{1 \leq i < j \leq L} |\omega_i^2 - \omega_j^2|^\beta \prod_{i=1}^L |\omega_i|^\alpha e^{-L \frac{\beta}{2} V(\omega_i^2)} d\omega_i$$

where $V(x), \alpha, \beta$ are parameters. The distribution makes sense for continuous values of α, β , but for special values of these parameters, it has an interpretation in terms of random matrix theory, as reviewed in section 2.1.

As shown in [66–68] and discussed in section 2.6 above, the symmetry classes associated to these ensembles are relevant the supercharge Q in the $\mathcal{N} = 1$ supersymmetric SYK model [81]. An important detail is that the eigenvalues of Q come in pairs $\omega, -\omega$. Also, for $\beta = 4$, there is twofold degeneracy. We emphasize that the probability distribution (4.49) is written for a set of independent variables, meaning just one of each pair of distinct eigenvalues $\omega, -\omega$. We will deal with the degeneracy later when we compare to JT supergravity.

In supersymmetric quantum mechanics, the Hamiltonian is the square of the supercharge $H = Q^2$, so eigenvalues of the Hamiltonian are $\lambda = \omega^2$. The measure (4.49) for the independent eigenvalues of Q implies a measure for the distinct eigenvalues of H :

$$(4.50) \quad \prod_{1 \leq i < j \leq L} |\lambda_i - \lambda_j|^\beta \prod_{i=1}^L |\lambda_i|^{\frac{\alpha-1}{2}} e^{-L \frac{\beta}{2} V(\lambda)} d\lambda_i, \quad \lambda_i \geq 0.$$

It will be slightly easier to derive the loop equations using the variables ω . However, because we will ultimately be interested in viewing $\lambda_i = \omega_i^2$ as the physical eigenvalues, we work in terms of resolvents defined as

$$(4.51) \quad R(x) = \sum_{i=1}^L \frac{1}{x - \omega_i^2}.$$

⁵²At least one of these ensembles, namely the bifundamental of $U(L) \times U(L)$, was studied in early literature on matrix models and two-dimensional gravity [98–100].

As before, we use the notation $R(x_1, \dots, x_n) = R(x_1) \dots R(x_n)$, and $I = \{x_1, \dots, x_n\}$. Connected correlators of these resolvents still have a $1/L$ expansion of the same form (4.6), and our goal will be to find a recursion for the $R_g(I)$ using the loop equations.

The loop equations follow from the statement

$$(4.52) \quad 0 = \frac{1}{\beta x} \int_{-\infty}^{\infty} d^L \omega \frac{\partial}{\partial \omega_a} \left[\frac{\omega_a}{x - \omega_a^2} R(I) \prod_{i < j} |\omega_i^2 - \omega_j^2|^\beta \prod_{i=1}^L |\omega_i|^\alpha e^{-L \frac{\beta}{2} V(\omega_i^2)} \right].$$

Distributing the derivative and simplifying using analogs of the three steps in (4.12), (4.13), (4.17), one finds

$$(4.53) \quad \left\langle \left(\left(1 - \frac{2}{\beta} \right) \partial_x + \frac{\alpha - 1}{\beta x} - LV'(x) \right) R(x, I) + R(x, x, I) + \frac{2}{\beta x} \sum_{k=1}^n \partial_{x_k} \frac{xR(x, I \setminus x_k) - x_k R(I)}{x - x_k} \right\rangle = -\frac{L}{x} \langle P(x; I) \rangle,$$

where now

$$(4.54) \quad P(x; I) = \sum_{a=1}^L \frac{xV'(x) - \omega_a^2 V'(\omega_a^2)}{x - \omega_a^2} R(I).$$

Again, $P(x; I)$ is analytic in x assuming an analytic potential. This can be rewritten in terms of connected correlators using the same logic as in the β -ensemble case. One finds

$$(4.55) \quad \left(\left(1 - \frac{2}{\beta} \right) \partial_x + \frac{\alpha - 1}{\beta x} - LV'(x) \right) \langle R(x, I) \rangle_c + \langle R(x, x, I) \rangle_c + \sum_{J \subseteq I} \langle R(x, J) \rangle_c \langle R(x, I \setminus J) \rangle_c + \frac{2}{\beta x} \sum_{k=1}^n \partial_{x_k} \frac{x \langle R(x, I \setminus x_k) \rangle_c - x_k \langle R(I) \rangle_c}{x - x_k} = -\frac{L}{x} \langle P(x; I) \rangle_c.$$

We will use this equation to compute the $R_g(I)$ following our steps for the β -ensemble. First, we will work out three special cases. Then we will write a recursion relation for the other cases.

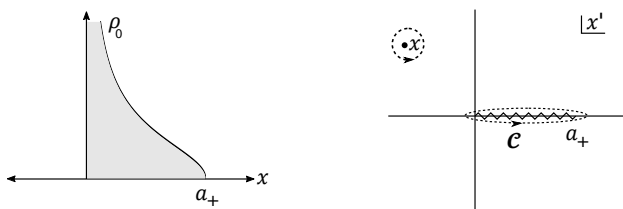


Figure 16: At left we plot a typical $\rho_0(x)$, the density of eigenvalues for $H = Q^2$. Note the $1/\sqrt{x}$ divergence at the origin, and the \sqrt{x} vanishing at $x = x_+$. At right we show the cut x' plane and the contour manipulation used in deriving the dispersion relation.

4.2.1. A first special case $R_0(x)$. As in the β -ensemble case, we start with the special case $I = \emptyset$, and further simplify by keeping only the terms proportional to L^2 . Then (4.55) reduces to

$$(4.56) \quad R_0(x)^2 - V'(x)R_0(x) = \frac{1}{x}(\text{analytic in } x).$$

which can be written as

$$(4.57) \quad y(x)^2 = \frac{1}{x}(\text{analytic in } x), \quad R_0(x) = \frac{V'(x)}{2} + y(x).$$

Again, we find a hyperelliptic curve, and the relationship between $y(x)$ and $\rho_0(x)$ in (4.24) remains true, where $\rho_0(x)$ is interpreted as the density of eigenvalues $\lambda = \omega^2$. Since these are non-negative, the density $\rho_0(x)$ has to be supported for $x \geq 0$. We will specialize to the simplest “one-cut” case, where $\rho_0(x)$ is nonzero in a single interval between zero and some value $a_+ > 0$. This interval will be referred to as “the cut.”

The important difference between the present case and the β -ensembles is the factor of $1/x$ in (4.57). What this means is that $y(x)$ will have a $1/\sqrt{x}$ singularity at the origin, and an ordinary \sqrt{x} branchpoint at a_+ . As an example, we can work out $y(x)$ for the case $V(x) = x$. Then (4.57), together with the requirement that $R_0(x) \approx 1/x$ for large x , implies that

$$(4.58) \quad y(x) = -\frac{1}{2}\sqrt{\frac{x-4}{x}} \quad \implies \quad \rho_0(x) = \frac{1}{2\pi}\sqrt{\frac{4-x}{x}}.$$

This is a special case of the Marchenko-Pastur distribution. It is the analog of the Wigner semicircle distribution for Wishart matrices $H = Q^2$. See figure 16.

The $1/\sqrt{x}$ singularity in y , compared to the typical behavior $y \sim (x - a)^{1/2}$ near a branch point in the case of a Dyson ensemble, reflects the fact that the eigenvalues $\lambda_i = \omega_i^2$ are strictly nonnegative in an Altland-Zirnbauer ensemble. This constraint leads to a divergence in the eigenvalue density at $x = 0$. Thus regardless of the values of α, β , an Altland-Zirnbauer ensemble is never equivalent to a Dyson ensemble, even though if one ignores the constraint $\lambda_i \geq 0$, the measure (4.50) in the special case $\alpha = 1$ would coincide with that of a Dyson ensemble.

4.2.2. A second special case $R_0(x, x_1)$. Next, take the case $I = \{x_1\}$ and study (4.55) at leading order L^1 . This leads to

$$(4.59) \quad -V'(x)R_0(x, x_1) + 2R_0(x)R_0(x, x_1) + \frac{2}{\beta x} \partial_{x_1} \frac{xR_0(x) - x_1R_0(x_1)}{x - x_1} = \frac{1}{x}(\text{analytic in } x).$$

Using (4.57), and moving some terms to the RHS, this can be rewritten as

$$(4.60) \quad 2y(x)R_0(x, x_1) + \frac{2}{\beta} \frac{y(x)}{(x - x_1)^2} = \frac{1}{x}(\text{analytic in } x \text{ in a neighborhood of the cut}).$$

We assume that x_1 is located away from the cut, so expressions like $1/(x - x_1)^p$ are analytic in x in a neighborhood of the cut. After multiplying through by x , this is

$$(4.61) \quad xy(x)R_0(x, x_1) + \frac{x}{\beta} \frac{y(x)}{(x - x_1)^2} \sim 0.$$

where as before, the \sim means equal up to terms that are analytic in x in a neighborhood of the cut. To analyze this equation, we specialize $\sigma(x)$ from (4.29) to the case with $a_- = 0$,

$$(4.62) \quad \sigma(x) = x(x - a_+),$$

and then multiply (4.60) by $\sqrt{\sigma(x)}/(xy(x))$. This ratio is analytic in x in a neighborhood of the cut, thanks to the fact that $y(x)$ has a $1/\sqrt{x}$ singularity

at $x = 0$, and a \sqrt{x} branch point at $x = a_+$. The resulting equation

$$(4.63) \quad R_0(x, x_1)\sqrt{\sigma(x)} + \frac{2}{\beta} \frac{\sqrt{\sigma(x)}}{(x - x_1)^2} \sim 0$$

is the same as we had for the β -ensemble. So we conclude that $R_0(x_1, x_2)$ is still given by the formula (4.32), but now with $a_- = 0$.

4.2.3. A third special case $R_0(x, x_1, x_2)$. As a final special case, we take $I = \{x_1, x_2\}$ and again work at leading order in $1/L$, which is now L^0 . Then (4.55) implies

$$(4.64) \quad xy(x)R_0(x, x_1, x_2) + xR_0(x, x_1)R_0(x, x_2) + \frac{x}{\beta} \left(\frac{R_0(x, x_2)}{(x - x_1)^2} + \frac{R_0(x, x_1)}{(x - x_2)^2} \right) \sim 0.$$

where the \sim means up to terms that are analytic in x in a neighborhood of the cut. After multiplying through by $\sqrt{\sigma(x)}/(xy(x))$, one can write exactly the same dispersion relation as (4.36). A difference is that in the present case, the factor $\sqrt{\sigma(x)}/y(x)$ has a zero at $x = 0$ which cancels the pole from the rest of the integrand in (4.36). This means that $R_0(x, x_1, x_2)$ receives a contribution only from the a_+ endpoint. This implies that $R_0(x_1, x_2, x_3)$ will vanish in a double-scaled (α, β) ensemble, where we take a_+ to infinity. In turn, this implies that all correlators $R_0(x_1, \dots, x_n)$ with $g = 0$ and $n \geq 3$ vanish in a double-scaled ensemble of this type.

4.2.4. The generic case. To consider the generic case, one plugs the general form of the genus expansion (4.6) into (4.55) and collects terms at order $L^{2-2g-|I|}$. The result can be written

$$(4.65) \quad 2xy(x)R_g(x, I) + xF_g(x, I) \sim 0$$

where

$$(4.66) \quad \begin{aligned} F_g(x, I) = & \left(\left(1 - \frac{2}{\beta}\right)\partial_x + \frac{\alpha-1}{\beta x} \right) R_{g-\frac{1}{2}}(x, I) + R_{g-1}(x, x, I) \\ & + \sum_{\text{stable}} R_h(x, J)R_{g-h}(x, I \setminus J) \\ & + 2 \sum_{k=1}^n \left(R_0(x, x_k) + \frac{1}{\beta} \frac{1}{(x-x_k)^2} \right) R_g(x, I \setminus x_k) \end{aligned}$$

The resulting dispersion relation is

$$(4.67) \quad R_g(x, I)\sqrt{\sigma(x)} = \frac{1}{2\pi i} \int_{\mathcal{C}} \frac{dx'}{x' - x} F_g(x', I) \frac{\sqrt{\sigma(x')}}{2y(x')}.$$

We will have more to say about special cases of these equations below. For now, we note that in the special case $(\alpha, \beta) = (1, 2)$, the first term in (4.66) vanishes. This implies by induction that all half-integer g coefficients vanish, and the dispersion relation reduces to a sum of residues at the endpoints. The argument for this is the same as for ordinary $\beta = 2$ ensembles in section 4.1.5.

Also, one can show that the ensembles with (α, β) equal to $(\alpha, 1)$ and $(-2\alpha + 3, 4)$ are related in the same way as discussed for the $\beta = 1$ and $\beta = 4$ ensembles in section 4.1.6. After rescaling the $(-2\alpha + 3, 4)$ resolvents as in (4.45), the results differ from the $(\alpha, 1)$ case only by the sign of the terms with half-integer g . Looking back to Table 4 of section 2.6.2, we deduce in the context of JT supergravity⁵³ that if N is even, then a shift $N \rightarrow N + 4$ has the effect only of changing the sign of crosscap contributions. The same is true for odd N . For $N = 3, 7$, the conclusion follows from the behavior under an exchange $(\alpha, 2) \leftrightarrow (2 - \alpha, 2)$. For $N = 1, 5$, one uses the original statement for $\beta = 1$ and $\beta = 4$ Dyson ensembles. See also appendix B.

4.2.5. The crosscap $R_{\frac{1}{2}}(x)$. For comparison to JT supergravity in section 5.3, it will be helpful to have a formula for the one-crosscap contribution to the resolvent, $R_{\frac{1}{2}}(x)$. This is determined by setting $g = \frac{1}{2}$ and $I = \emptyset$ in (4.66). The resulting equation can be written as

$$(4.68) \quad 2xy(x)R_{\frac{1}{2}}(x) + x \left(\left(1 - \frac{2}{\beta}\right)\partial_x + \frac{\alpha-1}{\beta x} \right) y(x) \sim 0.$$

This leads to

$$(4.69) \quad R_{\frac{1}{2}}(x) = \frac{1}{2\pi i} \int_{\mathcal{C}} \frac{dx'}{x' - x} \frac{\sqrt{\sigma(x')}}{\sqrt{\sigma(x)}} \left(\frac{\left(1 - \frac{2}{\beta}\right)y'(x')}{2y(x')} + \frac{\alpha-1}{2\beta x'} \right)$$

4.2.6. Double-scaling. An important detail is that JT gravity or supergravity is related not to an ordinary matrix integral of the type we have discussed here, but instead a particular type of limit that is referred to as a double-scaled matrix integral. We will briefly describe what this means.

⁵³We state the following for $\widehat{q} = 1 \pmod 4$, as assumed in the table. For $\widehat{q} = 3 \pmod 4$, one must exchange $N \rightarrow 8 - N$ in all statements.

In terms of the data of the spectral curve, to study a double-scaled matrix integral we scale the upper endpoint of the cut to infinity $a_+ \rightarrow \infty$, and we relax the constraint that $\rho_0(x)$ should have unit integral. Such a system can be understood as a limit of an ordinary matrix integral in which L is taken to infinity holding fixed the classical approximation to the density of eigenvalues near the lower endpoint. In an ordinary matrix integral, the classical approximation to the density of eigenvalues is $\rho_0^{\text{total}}(x) = L\rho_0(x)$, where ρ_0 is normalized so that the integral is one. To define a double-scaled integral, we write

$$(4.70) \quad \rho_0^{\text{total}}(x) = e^{S_0} \rho_0(x).$$

Then we take L to infinity and adjust the potential so that e^{S_0} remains fixed and $\rho_0(x)$ approaches the desired function.

We retain the definition $y(x + i\epsilon) = -i\pi\rho_0(x)$. Since the loop equations depend only on $y(x)$, they commute with the double-scaled limit. However, because L has been replaced by e^{S_0} in (4.70), the genus expansion is now in powers of e^{-S_0} :

$$(4.71) \quad \langle R(I) \rangle \simeq \sum_{g=0, \frac{1}{2}, 1, \dots} \frac{R_g(I)}{(e^{S_0})^{2g+|I|-2}}.$$

5. Comparison to JT gravity and supergravity

In this section, we will compare JT gravity and JT supergravity to the predictions of the loop equations for the matrix ensembles described in section 2.1. For ordinary JT gravity, there are three basic cases that correspond to the Dyson ensembles. As described in detail in sections 2.4 and 2.5, ordinary JT gravity theories with spin or pin structure sums reduce to these cases and need not be discussed further. On the other hand, there are ten versions of super JT gravity, covering both the Dyson and the Altland-Zirnbauer ensembles.

We will go through the cases and compare what we can between JT (super) gravity and the predictions of the loop equations. For quantities where we can compute the predictions of JT gravity and supergravity, we will find agreement with the loop equations. For other quantities, we will find predictions from the loop equations for the volumes of supermoduli space.

The basic dictionary that we will use to compare matrix integral quantities and JT gravity quantities is

$$(5.1) \quad \left\langle \text{Tr}(e^{-\beta_1 H}) \dots \text{Tr}(e^{-\beta_n H}) \right\rangle_{\text{matrix int.}}$$

(super) JT path integral summed over
topologies, with n asymptotic “trumpet” boundaries of regularized lengths
 \leftrightarrow
 β_j

On the LHS, H is the random matrix from an appropriate ensemble, including degeneracy if there is any.⁵⁴ On the RHS, we have a path integral in JT gravity or super gravity, with an asymptotic boundary of regularized length β for each insertion of $\text{Tr} e^{-\beta H}$ on the LHS.

In (5.1), we are studying the full (connected plus disconnected) correlation function on the LHS. And on the RHS, we have a sum over all bulk topologies, including disconnected ones. But to establish (5.1), it is sufficient to compare the connected correlators (or “cumulants”) constructed from both sides of (5.1). On the LHS this means that we subtract products of lower-order correlators. On the RHS, it means that we consider only connected geometries. After making this restriction, both sides have a “genus” expansion in powers of e^{-S_0}

$$(5.2) \quad \sum_{g=0}^{\infty} \frac{Z_g(\beta_1, \dots, \beta_n)}{(e^{S_0})^{2g+n-2}},$$

and we will compare the coefficients in the expansion $Z_g(\beta_1, \dots, \beta_n)$. In general, the sum will involve both half-integer and integer values of the “genus” summation index g .

On the matrix integral side, this genus expansion is determined by the loop equations. More precisely, in our discussion of the loop equations, we saw how to compute the e^{-S_0} expansion of correlation functions of resolvents, but one can go back and forth between resolvents and partition functions using Laplace and inverse Laplace transforms.

On the JT gravity side, $Z_g(\beta_1, \dots, \beta_n)$ is the path integral over two-dimensional connected geometries with fixed genus g . The weighting by powers of e^{S_0} comes from an Euler characteristic term we can add to the bulk

⁵⁴In cases where the relevant ensemble corresponds to a theory with an odd number of fermions, then we should multiply each trace in the LHS by $\sqrt{2}$, as explained in (2.7). For the most part we will consider NS boundaries, but we will comment on R boundaries in section 5.4.

JT action, as is familiar in string theory. The rest of the bulk action can be written as a topological BF theory, with gauge group $G = \text{SL}(2, \mathbb{R})$ for ordinary JT gravity and $G = \text{OSp}'(1|2)$ for JT supergravity. As described in section 3, the path integral of a BF theory reduces to an integral over the moduli space of flat connections, with measure given by the torsion.⁵⁵ So, the main part of the computation of $Z_g(\beta_1, \dots, \beta_n)$ is a computation of the volume of a moduli space with this measure.

More precisely, the path integral reduces to an integral over such a moduli space, together with a path integral over “wiggles” associated to the n asymptotic boundaries:

$$(5.3) \quad Z_g(\beta_1, \dots, \beta_n) = \int d(\text{moduli}) \int \mathcal{D}(\text{wiggles}) e^{-I_{\text{sch}}}.$$

These wiggles are described by the Schwarzian theory in the ordinary JT gravity case [4–6], and by the super-Schwarzian theory in the super JT gravity case [30]. These path integrals can be done in advance, and then glued together with the computation of the moduli space volume. For more on this, see appendix C.

Two different Schwarzian and super-Schwarzian path integrals will be needed, see figure 17. The first path integral, $Z^D(\beta)$ is an integral over the wiggles at the boundary of a disk. In this situation, the only parameter is the regularized length of the wiggly boundary, β . The second path integral, $Z^T(\beta, b)$, is an integral over the wiggles at the “big end” of a trumpet. The trumpet is defined so that the “small end” is a geodesic of length b , and the answer depends on this parameter as well as β , the regularized length of the big end. For both cases, the path integrals depend on the normalization of the path integral measure, and also on the normalization of β . Both of these normalizations are arbitrary, but must be chosen consistently. With a particular choice that we will describe in more detail in appendix C, the Schwarzian and super-Schwarzian path integrals for the disk and the trumpet are explicitly

$$(5.4) \quad Z_{\text{JT}}^D(\beta) = \frac{1}{4\pi^{1/2}\beta^{3/2}} e^{\frac{\pi^2}{\beta}} \quad Z_{\text{JT}}^T(\beta, b) = \frac{1}{2\sqrt{\pi\beta}} e^{-\frac{b^2}{4\beta}}$$

$$(5.5) \quad Z_{\text{SJT}}^D(\beta) = \sqrt{\frac{2}{\pi\beta}} e^{\frac{\pi^2}{\beta}} \quad Z_{\text{SJT}}^T(\beta, b) = \frac{1}{\sqrt{2\pi\beta}} e^{-\frac{b^2}{4\beta}}.$$

⁵⁵One also has to restrict to the appropriate topological component of the moduli space, and quotient by the action of the mapping class group.

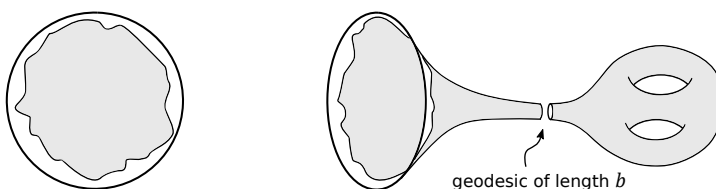


Figure 17: The disk (left) and trumpet (right) partition functions are path integrals over a wiggly regularized AdS_2 boundary, together with a fermionic partner mode in the super case. The length of the wiggly boundary is proportional to β . In the case of the trumpet, the other end of the trumpet is a geodesic of length b , across which the geometry can be glued to a bordered (super) Riemann surface.

The ordinary JT cases were computed in [12–20] and the super cases are computed in appendix C.

For the disk topology, there is no moduli space to integrate over, so the path integral over the Schwarzian “wiggles” is actually the whole answer:

$$(5.6) \quad Z_0(\beta) = Z_{(\text{S})\text{JT}}^D(\beta).$$

In comparing to matrix integrals, the disk plays a special role, because it determines the spectral curve, or equivalently the $\rho_0(x)$ function. We can read this off by interpreting $Z_0(\beta) = \int_0^\infty dx e^{-\beta x} \rho_0(x)$ and solving for $\rho_0(x)$. For the two cases of JT gravity and super JT gravity, one finds the $\rho_0(x)$ functions and corresponding spectral curves

$$(5.7) \quad \rho_{0\text{JT}}(x) = \frac{\sinh(2\pi\sqrt{x})}{4\pi^2} \quad y_{\text{JT}}(x) = \frac{\sin(2\pi\sqrt{-x})}{4\pi}$$

$$(5.8) \quad \rho_{0\text{SJT}}(x) = \frac{\sqrt{2} \cosh(2\pi\sqrt{x})}{\pi\sqrt{x}} \quad y_{\text{SJT}}(x) = -\frac{\sqrt{2} \cos(2\pi\sqrt{-x})}{\sqrt{-x}}.$$

As we have seen, together with the discrete choice of matrix ensemble, the data of the spectral curve completely determines the genus expansion of a matrix integral. So from the perspective of a correspondence between (super) JT gravity and matrix integrals, the disk topology should be regarded as input that allows us to fit the correct matrix integral. All other topologies constitute nontrivial tests of the correspondence.

In (super) JT gravity, the path integrals on other topologies Z_g are given by a combination of the Schwarzian trumpet path integral and the volumes of moduli space. There are two cases that have to be treated individually.

One of them is the “double trumpet,” for which the path integral gives

$$(5.9) \quad Z_0(\beta_1, \beta_2) = c \int_0^\infty b \, db \, Z_{(S)JT}^T(\beta_1, b) Z_{(S)JT}^T(\beta_2, b).$$

In this expression, two trumpets have been glued together by integrating over the length of their shared geodesic b and the relative twist. The integral over the relative twist gives the factor of b in the measure. In a theory where we sum over neither spin structures nor orientation reversal, the constant c should be one. In a theory where we sum over one but not the other, we have $c = 2$. Finally, in a theory where we sum over both spin structures and orientation reversal, we have $c = 4$.

Another special case is the crosscap spacetime:

$$(5.10) \quad Z_{\frac{1}{2}}(\beta) = \int_0^\infty db \, V_{\frac{1}{2}}(b) Z_{(S)JT}^T(\beta, b).$$

This case is special because when we glue the trumpet to the crosscap, there is only a single modulus involved in the gluing, the size. There is no twist modulus because of the rotational symmetry of the crosscap. The result of this is that the measure factor is simply db rather than bdb as we have in other cases.

All other $Z_g(\beta_1, \dots, \beta_n)$ are given by the generic formula⁵⁶

$$(5.11) \quad Z_g(\beta_1, \dots, \beta_n) = c_n \int_0^\infty \prod_{j=1}^n [b_j \, db_j \, Z_{(S)JT}^T(\beta_j, b_j)] V_g(b_1, \dots, b_n).$$

The volumes $V_g(b_1, \dots, b_n)$ are computed using the torsion as described in section 3.4.7 and section 3.5.3. We define them to also include the sum over spin structures (if present) holding fixed the NS spin structure on the boundaries.⁵⁷ The constant c_n should be one in a theory where we do not gauge orientation reversal, and it should be 2^{n-1} in a theory where we do. This accounts for the possibility of gluing the n trumpets in with independent orientation-reversals, up to an overall orientation reversal.

⁵⁶In principle, the genus one case with a single boundary is also special: because the torus with one hole has a \mathbb{Z}_2 symmetry, we should integrate the twist from zero to $b/2$ instead of b . Instead, we simply define $V_1(b)$ to be one-half of the true moduli space volume of the torus with one hole, and integrate the twist as usual. See also footnote 72.

⁵⁷We will consider Ramond boundaries briefly in section 5.4.

In order to compare to the loop equations, we will often work in terms of resolvents instead of partition functions. These can be obtained using $R(x) = -\int_0^\infty d\beta e^{\beta x} \text{Tr}(e^{-\beta H})$. In order to apply this integral transform to the (super) JT partition functions, we will often use the formulas

$$(5.12) \quad \begin{aligned} \int_0^\infty d\beta e^{-\beta z^2} Z_{\text{JT}}^T(\beta, b) &= \frac{1}{2z} e^{-bz}, \\ \int_0^\infty d\beta e^{-\beta z^2} Z_{\text{SJT}}^T(\beta, b) &= \frac{1}{\sqrt{2}z} e^{-bz}. \end{aligned}$$

5.1. Ordinary JT gravity with time-reversal symmetry

JT gravity on orientable surfaces was related to a double-scaled $\beta = 2$ matrix ensemble in [35]. In this section we will relate $\beta = 1, 4$ versions of the same matrix ensemble to a JT gravity theory in which orientation-reversal is viewed as a gauge symmetry.

Concretely, gauging orientation-reversal means we can glue surfaces together with an orientation reversal, so in particular we should include unorientable surfaces as well as orientable ones. There are two versions of the bulk theory, depending on whether or not we include a factor of $(-1)^{n_c}$ where n_c is the number of crosscaps. These two versions are related to $\beta = 1$ and $\beta = 4$ ensembles. Because of the discussion in 4.1.6, it will be sufficient to compare just one of these cases, and we choose the $\beta = 1$ ensemble, which is dual to a bulk theory in which we do not include $(-1)^{n_c}$.

To compare the matrix and JT gravity predictions, we will start with genus zero. Here we can use as input the fact from [35] that JT gravity is dual to a $\beta = 2$ ensemble. It is easy to check from the recursion that genus zero resolvents in the $\beta = 1$ and $\beta = 2$ ensemble are related by the simple factor

$$(5.13) \quad R_0^{\beta=1}(I) = 2^{n-1} R_0^{\beta=2}(I), \quad n = |I|.$$

We wrote this equation in terms of resolvents, but partition functions $Z_0(\beta_1, \dots, \beta_n)$ will also be related by the same factor of 2^{n-1} . So, to show agreement with the matrix integral, we need to find the same factor in the relationship between JT gravity partition functions in the theories with and without orientation-reversal gauged. Indeed, this is just the factor c_n discussed in (5.11). Topologically, $Z_0(\beta_1, \dots, \beta_n)$ involves n trumpets glued into a sphere. We can glue each trumpet to the sphere with or without an orientation reversal, giving a factor of 2^n . Gauging the overall orientation gives a single factor of $1/2$, so $c_n = 2^{n-1}$, which matches (5.13).

Next we discuss genus one-half with a single boundary. From the loop equations for the ordinary β -ensembles, the equation (4.37) for $R_{\frac{1}{2}}(x)$ is

$$(5.14) \quad 2y(x)R_{\frac{1}{2}}(x) + \left(1 - \frac{2}{\beta}\right) \partial_x y(x) \sim 0.$$

where we replaced $\partial_x R_0(x)$ by $\partial_x y(x)$ at the cost of introducing an analytic term that we ignore in this expression. The corresponding dispersion relation is

$$(5.15) \quad R_{\frac{1}{2}}(x)\sqrt{\sigma(x)} = \frac{1 - \frac{2}{\beta}}{2\pi i} \int_{\mathcal{C}} \frac{dx'}{x' - x} \sqrt{\sigma(x')} \frac{\partial_{x'} y(x')}{2y(x')}.$$

For a double-scaled matrix integral with $a_- = 0$ and $a_+ = \infty$, this becomes (for $\beta = 1$)

$$(5.16) \quad \sqrt{-x}R_{\frac{1}{2}}(x) = -\frac{1}{2\pi} \int_0^\infty \frac{\sqrt{x'} dx'}{x' - x} \frac{\partial_{x'} y(x')}{y(x')}.$$

For a spectral curve $y(x)$ that is asymptotically a power of x , this integral would converge. But for the $y(x)$ in (5.7), it is logarithmically divergent at large x' . What this means is that the limiting procedure needed to define the double-scaled $\beta = 1$ theory for JT gravity doesn't produce a finite e^{-S_0} expansion. (The divergence is regulated in the minimal string, see appendix F.)

For our purposes, this is actually a good thing, because as we saw in section 3.4.6, the volume of the moduli space of unorientable surfaces is in general divergent. For the quantity $R_{\frac{1}{2}}$, we will be able to compare the integrands of the matrix and JT expressions, and find precise agreement. To proceed, we substitute in (5.7) and rewrite (5.16) as

$$(5.17) \quad R_{\frac{1}{2}\text{M}}(-z^2) = -\frac{1}{2z} \int_0^\infty \frac{dx'}{x' + z^2} \coth(2\pi\sqrt{x'}).$$

Here we added a subscript “M” to indicate that it is the matrix integral answer. One would like to use this formula to predict an answer for the path integral measure for the crosscap, and then check it against the torsion computation (3.56). If we didn't know the measure for the crosscap in JT gravity, we would write (5.10) with $V_{\frac{1}{2}}(b)$ unknown. This can be converted to an expression for the resolvent:

$$(5.18) \quad R_{\frac{1}{2}\text{JT}}(-z^2) = -\int_0^\infty d\beta e^{-\beta z^2} Z_{\frac{1}{2}\text{JT}}(\beta) = -\frac{1}{2z} \int_0^\infty db e^{-az} V_{\frac{1}{2}}(b).$$

In the last step we inserted (5.10) and did the integral over β explicitly using (5.12). So, matching (5.17) and (5.18), the equation to be solved for $V_{\frac{1}{2}}(b)$ is

$$(5.19) \quad \int_0^\infty db e^{-bz} V_{\frac{1}{2}}(b) = \int_0^\infty \frac{dx'}{x' + z^2} \coth(2\pi\sqrt{x'}).$$

Taking an inverse Laplace transform of both sides of the equation, one finds

$$(5.20) \quad V_{\frac{1}{2}}(b) = \int \frac{dz}{2\pi i} e^{bz} \int_0^\infty \frac{dx'}{x' + z^2} \coth(2\pi\sqrt{x'})$$

$$(5.21) \quad = \int_0^\infty \frac{dx'}{\sqrt{x'}} \sin(b\sqrt{x'}) \coth(2\pi\sqrt{x'})$$

$$(5.22) \quad = \frac{1}{2} \coth\left(\frac{b}{4}\right).$$

This matches the answer we got from the computation of the torsion of the crosscap (3.56).

It may be possible to make further comparisons between these two divergent theories, but we will stop at this point and move to super JT gravity.

5.2. Super JT without time-reversal symmetry

We will begin by considering a version of JT supergravity in which we do not gauge orientation-reversal. This means that we are restricted in how surfaces can be glued together. In particular, the twisted double-trumpet and its relatives are not allowed, and nor are unorientable surfaces. This is expected to be dual to a matrix ensemble that does not respect any time-reversal symmetry. As described in section 2.6 and table 3, there are two versions of the matrix ensemble, and two versions of the bulk theory.

In this section and in the next one, we will focus on the case that the boundaries have NS spin structures. We will comment on the case with R boundaries in section 5.4.

5.2.1. Sum of even and odd spin structures. We start with the case corresponding to odd N . This is a theory with an anomaly in $(-1)^F$, which (as described in section 2.6.3) is already accounted for in the contribution of the boundary fermion in the Schwarzian supermultiplet. So in the bulk theory we should not include the $(-1)^\zeta$ TFT. This means that bulk spin structures are all weighted equally, with no minus signs. As argued in section 2.6, the random matrix ensemble is specified by saying that the supercharge Q is drawn from an ordinary $\beta = 2$ ensemble.

Since we are in an ensemble with an odd number of fermions, the relationship between the path integral and the trace in the Hilbert space involves a factor of $\sqrt{2}$, see (2.7). So we expect to have the relationship

$$(5.23) \quad Z_{\text{SJT}}(\beta) = \sqrt{2} \text{Tr}(e^{-\beta Q^2}).$$

To determine the spectral curve of the matrix ensemble, we need the leading density of eigenvalues of Q . This can be obtained by taking the genus zero approximation of both sides of (5.23):

$$(5.24) \quad Z_{\text{SJT}}^D(\beta) = \sqrt{2} \int_{-\infty}^{\infty} dx \rho_0(x) e^{-\beta x^2}.$$

After plugging in (5.5) for the LHS, we need to solve this equation for $\rho_0(x)$. It is easy to check the solution

$$(5.25) \quad \rho_0(x) = \frac{\cosh(2\pi x)}{\pi}.$$

This is the leading density of eigenvalues for the matrix Q . The corresponding spectral curve is determined by (4.24). It is somewhat unusual, with a branch cut running along the whole real axis. This can be arranged as a limit of a conventional matrix integral where both endpoints a_{\pm} have been scaled away, to plus and minus infinity.

We can get the expression for $R_0(x_1, x_2)$ for such a matrix integral by taking $a_{\pm} = \pm\infty$ in (4.32). The result for $\beta = 2$ is

$$(5.26) \quad R_0(x_1, x_2) = \begin{cases} 0 & x_1, x_2 \text{ on same side of cut} \\ -\frac{1}{(x_1 - x_2)^2} & x_1, x_2 \text{ on opposite sides.} \end{cases}$$

Here the cut runs along the whole real axis, so the statement is that R_0 is nonzero (on the principal sheet) only if the two arguments are in opposite half-planes. Eq. (5.26) gives the two-resolvent correlator for resolvents constructed from the eigenvalues of Q . To compare to JT supergravity, it is helpful to first convert to a formula for the resolvents of H . Using

$$(5.27) \quad \frac{1}{x + \lambda_Q} - \frac{1}{x - \lambda_Q} = \frac{2x}{x^2 - \lambda_Q^2} = \frac{2x}{x^2 - \lambda_H}$$

we can relate the resolvents (using an obvious notation) as

$$(5.28) \quad R^{(Q)}(x) - R^{(Q)}(-x) = 2xR^{(H)}(x^2).$$

The pair correlator of the resolvents of H is determined by this formula and the pair correlator of resolvents of Q in (5.26). The result is

$$(5.29) \quad R_0^{(H)}(-z_1^2, -z_2^2) = \frac{1}{2z_1z_2(z_1 + z_2)^2}.$$

Including the factor of $\sqrt{2}$ in (5.23), the matrix integral result will be consistent with JT supergravity if $R_{0\text{SJT}}$ is equal to $(\sqrt{2})^2 = 2$ times this value. Let's now check against the prediction from JT supergravity. From (5.9) and after plugging in (5.5), we have

$$(5.30) \quad Z_{0\text{SJT}}(\beta_1, \beta_2) = 2 \int_0^\infty b \, db \, Z_{\text{SJT}}^T(\beta_1, b) Z_{\text{SJT}}^T(\beta_2, b) = \frac{2\sqrt{\beta_1\beta_2}}{\pi(\beta_1 + \beta_2)}.$$

The factor of two out front corresponds to the value $c = 2$ in (5.9) which is appropriate for the current case where we sum over spin structures but not over orientation-reversal. Translating to resolvents using $R(x) = -\int_0^\infty d\beta e^{\beta x} Z(\beta)$ and (5.12), we indeed find twice the value in (5.29).

A property of this type of scaled matrix integral is that all higher-order correlators vanish. This is due to the following. In a $\beta = 2$ ensemble, as we remarked in section 4.1.5, the recursion reduces to a sum of residues at the endpoints of the cut. Because of the factor $1/(x' - x)$ in the integrand of the dispersion relation (4.41), these residues vanish in the limit that the endpoints go to infinity. So for a $\beta = 2$ matrix integral that has been double scaled to give a spectral curve like (5.25), all higher order correlators are zero.

The conjectured correspondence of JT supergravity with a matrix integral then predicts that the volumes of the moduli spaces of bordered super-Riemann surfaces (summing equally over even and odd spin structures) are zero. We confirm this at the end of appendix D.

5.2.2. Difference of even and odd spin structures. Now we consider the random matrix ensemble appropriate to the case with an even number of Majorana fermions. We expect this to be related to JT supergravity including a factor of $(-1)^\zeta$, which means that we take the sum of even spin structures minus the sum of odd ones. As argued in section 2.6, the appropriate random matrix ensemble is a $(\alpha, \beta) = (1, 2)$ ensemble. Because the Hamiltonian is two-fold degenerate, we expect

$$(5.31) \quad Z_{\text{SJT}}(\beta) = \text{Tr}(e^{-\beta H}) = 2 \sum_{\substack{\text{distinct} \\ \text{eigs. of } H}} e^{-\beta \lambda_i} = 2Z_{\text{M}}(\beta).$$

Here, we are using the subscript M to indicate a matrix integral quantity where we sum only over the distinct eigenvalues of H , as we did in our study of the loop equations above.

Because of the degeneracy, the density of distinct eigenvalues ρ_M is half of the total density of eigenvalues. For JT supergravity, the genus zero expression for the total density of eigenvalues is given in (5.8). So we learn that ρ_{0M} and the spectral curve y of the matrix integral are given by

$$(5.32) \quad \rho_{0M}(x) = \frac{\rho_{0\text{SJT}}(x)}{2} = \frac{\cosh(2\pi\sqrt{x})}{\pi\sqrt{2x}}, \quad y_M(x) = -\frac{\cos(2\pi\sqrt{-x})}{\sqrt{-2x}}.$$

Recall that for the (α, β) -ensembles, we work directly with resolvents for H , and not for Q . So here we have written the leading density of distinct eigenvalues of H .

We would now like to use the loop equations of the matrix integral to predict answers for JT supergravity. It is convenient to compare the resolvents in the two theories. Because of the degeneracy, multi-resolvents in JT supergravity will be related to multi-resolvents of the distinct eigenvalues as

$$(5.33) \quad R_{\text{SJT}}(I) = 2^{|I|} R_M(I).$$

As a first example, we can compare the leading two-point correlator. From the expression (4.34) specialized to $(\alpha, \beta) = (1, 2)$, one finds

$$(5.34) \quad R_{0M}(-z_1^2, -z_2^2) = \frac{1}{4z_1z_2(z_1 + z_2)^2}.$$

Because of (5.33), the prediction for JT gravity is four times this, which is indeed the answer that follows from (5.30).

To evaluate higher order correlators on the matrix side, we use the recursion relation (4.67). For a double-scaled theory of the type we are studying, it is convenient to write this dispersion relation using a uniformizing coordinate $z = \sqrt{-x}$ as

$$(5.35) \quad R_{gM}(-z^2, I) = \frac{1}{2\pi iz} \int_{\epsilon-i\infty}^{\epsilon+i\infty} \frac{z'^2 dz'}{z'^2 - z^2} \frac{F_{gM}(-z'^2, I)}{y_M(-z'^2)}.$$

In terms of z , the spectral curve (5.32) is

$$(5.36) \quad x = -z^2, \quad y_M = -\frac{\cos(2\pi z)}{\sqrt{2}z}.$$

The quantity F_g is given in (4.66). An important feature is that the first term in the definition of F_g vanishes for the ensemble $(\alpha, \beta) = (1, 2)$. An argument similar to the one in section 4.1.5 then shows that the dispersion relation (4.67) reduces to a sum over residues at the endpoints of the cut in the x plane. (For the double-scaled case (5.35), this is simply one-half the residue at $z' = 0$.) This implies that the loop equations reduce to “topological recursion” for the spectral curve (5.36).

Now, assuming a correspondence between JT supergravity and matrix integrals, this recursion can be used to predict the volumes of super moduli space. The idea is as follows. In JT supergravity, correlators of partition functions are related to volumes of super moduli space by (5.11). For the present case where we do not sum over orientation-reversal, we have $c_n = 1$. Using $R(x) = -\int_0^\infty d\beta e^{\beta x} Z(\beta)$, and using (5.12), we find the JT supergravity formula for the resolvents in terms of the volumes

$$(5.37) \quad R_{g\text{SJT}}(-z_1^2, \dots, -z_n^2) = (-1)^n \int_0^\infty V_g(b_1, \dots, b_n) \prod_{j=1}^n b_j db_j \frac{e^{-b_j z_j}}{\sqrt{2} z_j}.$$

In this expression, the $V_g(b_1, \dots, b_n)$ are the volumes of moduli space of bordered super Riemann surfaces, with sum over spin structures weighted by $(-1)^\zeta$.⁵⁸ An inverse expression gives the volumes in terms of the resolvents: (5.38)

$$V_g(b_1, \dots, b_n) = (-1)^n \int_{c_0+i\mathbb{R}} R_{g\text{SJT}}(-z_1^2, \dots, -z_n^2) \prod_{j=1}^n \frac{dz_j}{2\pi i} \frac{\sqrt{2} z_j}{b_j} e^{b_j z_j}.$$

Assuming the correspondence between SJT and a matrix integral (5.33) and using the matrix integral recursion (5.35), one can efficiently compute the $R_{g\text{SJT}}$ functions. Then one can compute the inverse Laplace transform (5.38) to obtain V_g . As an example, the following three cases can be computed (in order) using the recursion:

$$(5.39) \quad \begin{aligned} R_{1\text{SJT}}(-z_1^2) &= \frac{1}{2^{7/2} z_1^3} & R_{1\text{SJT}}(-z_1^2, -z_2^2) &= \frac{1}{2^4 z_1^3 z_2^3} \\ R_{2\text{SJT}}(-z_1^2) &= \frac{9}{2^{17/2} z_1^5} + \frac{9\pi^2}{2^{15/2} z_1^3}. \end{aligned}$$

⁵⁸Alternatively, since we know that the sum without this factor vanishes, we can regard it as twice the sum over even spin structures, or minus two times the sum over odd spin structures.

After using (5.38), one finds the corresponding volumes (see footnote 56)

$$(5.40) \quad V_1(b_1) = -\frac{1}{8} \quad V_1(b_1, b_2) = \frac{1}{8} \quad V_2(b_1) = -\frac{3}{2^9}(b_1^2 + 12\pi^2).$$

In practical terms, this is the most efficient method to compute the volumes. But it is also possible to rewrite the recursion (5.35) directly in terms of the volumes. This gives a recursion that is structurally identical to the one derived by Mirzakhani for the volumes of ordinary moduli spaces.

To derive this relation, one can substitute the recursion (5.35) into the RHS of (5.38) taking care of the factors of two due to (5.33). This gives an expression for $V_g(b_1, \dots, b_n)$ in terms of some integrals of resolvents. Then using (5.37) one rewrites these resolvents in terms of “lower order” volumes. The resulting expression contains many integrals, but most of them are Laplace and inverse Laplace transforms that cancel each other. What remains is a small number of z and b integrals. One of the z integrals can be done by trivial contour integration, and the other can be treated using the following function

$$(5.41) \quad \begin{aligned} D(x, y) &= \int_{-i\infty}^{i\infty} \frac{dz}{2\pi i} \frac{e^{-xz}}{\cos(2\pi z)} \sinh(yz) \\ &= \frac{1}{8\pi} \left(\frac{1}{\cosh \frac{x-y}{4}} - \frac{1}{\cosh \frac{x+y}{4}} \right). \end{aligned}$$

The final b integrals are explicit in the final form of the recursion relation, which is

$$(5.42) \quad \begin{aligned} bV_g(b, B) &= -\frac{1}{2} \int_0^\infty (b' db')(b'' db'') D(b' + b'', b) \\ &\quad \times \left(V_{g-1}(b', b'', B) + \sum_{\text{stable}} V_{h_1}(b', B_1) V_{h_2}(b'', B_2) \right) \\ &\quad - \sum_{k=1}^{|B|} \int_0^\infty b' db' \left(D(b' + b_k, b) + D(b' - b_k, b) \right) V_g(b', B \setminus b_k). \end{aligned}$$

In this expression, we are using the notation $B = \{b_1, \dots, b_n\}$. In the second term on the first line, the sum marked “stable” is over $h_1 + h_2 = g$ and $B_1 \cup B_2 = B$, with terms involving the unstable cases $V_0(b_i)$ and $V_0(b_i, b_j)$ excluded. We can start the recursion with $V_1(b_1) = -1/8$ along with the vanishing of the volumes in genus 0.

The volumes of supermoduli space, like the volumes of moduli space, depend on the normalization of the symplectic form. Rescaling this form will multiply the entire RHS of (5.42) by a constant factor. Formula (5.42) is supposed to be correct with the symplectic form normalized so that the bosonic part of it reduces to the form $\omega = \sum_i db_i d\tau_i$ in terms of Fenchel-Nielsen (length-twist) coordinates measured in units where the hyperbolic curvature length is one, so the scalar curvature is $R = -2$.

The recursion relation (5.42) has also been obtained in unpublished work by Norbury, extending work on the “cosine” spectral curve [56] (and earlier work on the related “Bessel” spectral curve [57]). Norbury has also computed examples such as those of eqn. (5.40).

In the case of ordinary moduli space, a recursion relation structurally similar to (5.42) was derived by Mirzakhani [37] using a sum rule for geodesics on hyperbolic surfaces. Shortly afterwards, Eynard and Orantin showed that Laplace transforms of Mirzakhani’s volumes satisfy topological recursion [36]. We have here obtained eqn. (5.42) purely as a statement about a matrix ensemble, without reference to super Riemann surfaces. However, in appendix D, we show how eqn. (5.42) can be obtained, as a statement about supermoduli space volumes, by a supersymmetric analog of Mirzakhani’s approach. This will establish the connection between super JT gravity and the random matrix ensemble, in the same sense that the relation between bosonic JT gravity and a matrix ensemble was shown in [35].

5.3. Super JT with time-reversal symmetry

Next we will study the correspondence between JT supergravity and matrix integrals in the case where we include time-reversal symmetry. In JT supergravity, this means that we will gauge orientation-reversal in the bulk. In section 3.5.1, it was shown that the bulk theory will involve a sum over pin^- structures, rather than pin^+ . In section 2.6 and table 4, we described the various matrix ensembles that appear, and the eight corresponding bulk topological field theories $e^{-i\pi\eta N'/2}$. We will now compare what we can between the matrix ensembles and JT theories, starting with genus zero.

5.3.1. Genus zero.

One boundary. At genus zero with one boundary, we do not have any prediction from the matrix integral. Instead, by matching to super JT gravity, we simply learn what the spectral curve of the matrix integral should be.

To get this right, we have to understand a normalization factor in the relationship between matrix integral quantities and JT gravity quantities. One nontrivial factor was noted in (2.7) above: in the case with an odd number of fermions, the path integral gives $\sqrt{2}$ times the trace in the Hilbert space. The other factor is due to the degeneracy of the levels of the Hamiltonian. It is convenient to combine these together into a variable γ

$$(5.43) \quad \gamma = \begin{cases} (\text{degeneracy in } H) & N \text{ even} \\ \sqrt{2} \cdot (\text{degeneracy in } H) & N \text{ odd} \end{cases}$$

so that the relationship between the super JT quantities and the matrix quantities is expected to be

$$(5.44) \quad R_{\text{SJT}}(I) = \gamma^{|I|} R_{\text{M}}(\beta).$$

The values of γ for the various cases are given in table 4. For the six cases corresponding to (α, β) ensembles, it is straightforward to obtain the spectral curve, generalizing the argument that led to (5.32) for the case $\gamma = 2$. The density of distinct eigenvalues of H is equal to $1/\gamma$ times the total density inferred from the JT supergravity path integral, and the spectral curve is therefore

$$(5.45) \quad y_{\text{M}}(x) = \frac{1}{\gamma} y_{\text{SJT}}(x), \quad y_{\text{SJT}}(x) = -\frac{\sqrt{2} \cos(2\pi\sqrt{-x})}{\sqrt{-x}}.$$

For the two β -ensemble cases, the relevant spectral curve is the one related to the density of eigenvalues of Q . In (5.25) we determined the density of distinct eigenvalues of Q for the case with $\gamma = \sqrt{2}$. Repeating the argument for a general value of γ , we find

$$(5.46) \quad \rho_0^{(Q)}(x) = \frac{\sqrt{2} \cosh(2\pi x)}{\gamma \pi}.$$

Two boundaries. At genus zero with two boundaries, the SJT gravity prediction is

$$(5.47) \quad Z_{0 \text{ SJT}}(\beta_1, \beta_2) = 2 \cdot 2 \int_0^\infty b \, db \, Z_{\text{SJT}}^T(\beta_1, b) Z_{\text{SJT}}^T(\beta_2, b) = \frac{4\sqrt{\beta_1\beta_2}}{\pi(\beta_1 + \beta_2)}.$$

The factors of two out front are because of the sum over pin^- structures and the sum over the double trumpet and twisted double trumpet. Using

$R(x) = -\int_0^\infty d\beta e^{\beta x} Z(\beta)$, this implies

$$(5.48) \quad R_{0\text{SJT}}(-z_1^2, -z_2^2) = \frac{2}{z_1 z_2 (z_1 + z_2)^2}.$$

We would like to compare this to the random matrix ensembles. For the six (α, β) cases in table 4, the leading two-resolvent correlator is given in (4.34). The x dependence is the same as in (5.48), but the prefactor is $\frac{1}{2\beta}$ instead of 2. However, to convert from $R_{0\text{M}}(x_1, x_2)$ to $R_{0\text{SJT}}(x_1, x_2)$, we need to multiply by γ^2 . For each of the six (α, β) cases in the table, one finds that $\gamma^2/(2\beta) = 2$, so we find agreement with (5.47).

For the remaining two β -ensemble cases from table 4, we need a separate argument. For these cases, $H = Q^2$, where Q is either a $\beta = 1$ or $\beta = 4$ matrix from a regular β -ensemble. Repeating the steps that led to (5.29) for a general value of β , one finds

$$(5.49) \quad R_{0\text{M}}(-z_1^2, -z_2^2) = \frac{1}{\beta z_1 z_2 (z_1 + z_2)^2}.$$

Multiplying by γ^2 , we again find agreement with (5.48) for the two β -ensemble cases in table 4.

More boundaries. As shown in appendix A, in super JT gravity, all higher correlators vanish at genus zero, because they are proportional to the volume of a symplectic supermanifold with more fermionic than bosonic coordinates. This is consistent with the matrix integral. For the six (α, β) cases, this follows from the discussion in section 4.2.3. For the two β -ensemble cases, one can argue as follows. At genus zero, we noted in our discussion of β -ensembles in section 4.1.3 that the dispersion relation for $R_0(x_1, x_2, x_3)$ reduces to a sum of residues at the endpoints of the cut. So we can apply the logic from the end of section 5.2.1: in the double-scaled limit that is applied to Q , both endpoints move to $\pm\infty$. Then the residues vanish due to the factor of $\frac{1}{x'-x}$, where x remains finite and x' is evaluated at one of the endpoints. It is easy to check from the recursion that vanishing of $R_0(x_1, x_2, x_3)$ implies that all higher $R_0(I)$ vanish.

5.3.2. Genus one-half. At genus one-half, we will get a more detailed check on the correspondence by computing $R_{\frac{1}{2}}(x)$. As before, we start with the six (α, β) ensemble cases. The matrix integral formula for $R_{\frac{1}{2}}(x)$ is (4.69).

Substituting in the spectral curve (5.45) and simplifying, one finds

$$(5.50) \quad R_{\frac{1}{2}M}(-z^2) = \frac{1}{2z} \int_0^\infty \frac{\sqrt{x'} dx'}{x' + z^2} \left[\left(1 - \frac{2}{\beta}\right) \frac{\tanh(2\pi\sqrt{x'})}{\sqrt{x'}} + \frac{\alpha - \frac{\beta}{2}}{\pi\beta x'} \right].$$

As in the bosonic case, let us pretend that we don't know the answer in super JT, and write

$$(5.51) \quad R_{\frac{1}{2}SJT}(x) = - \int_0^\infty d\beta e^{\beta x} Z_{\frac{1}{2}SJT}(x)$$

$$(5.52) \quad = - \int_0^\infty d\beta e^{\beta x} \int_0^\infty db Z_{SJT}^T(\beta, b) V_{\frac{1}{2}}(b)$$

with an undetermined $V_{\frac{1}{2}}(b)$. Evaluating this at $x = -z^2$ and doing the integral over β using (5.12), we find

$$(5.53) \quad R_{\frac{1}{2}SJT}(-z^2) = - \frac{1}{\sqrt{2}z} \int_0^\infty db e^{-bz} V_{\frac{1}{2}}(b).$$

Setting $R_{\frac{1}{2}SJT} = \gamma R_{\frac{1}{2}M}$, and taking the inverse Laplace transform with respect to z , we find

$$(5.54) \quad V_{\frac{1}{2}}(b) = - \frac{\gamma}{\sqrt{2}} \int \frac{dz}{2\pi i} e^{bz} \int_0^\infty \frac{\sqrt{x'} dx'}{x' + z^2} \left[\left(1 - \frac{2}{\beta}\right) \frac{\tanh(2\pi\sqrt{x'})}{\sqrt{x'}} + \frac{\alpha - \frac{\beta}{2}}{\pi\beta x'} \right]$$

$$(5.55) \quad = - \frac{\gamma}{\sqrt{2}} \int_0^\infty dx' \sin(b\sqrt{x'}) \left[\left(1 - \frac{2}{\beta}\right) \frac{\tanh(2\pi\sqrt{x'})}{\sqrt{x'}} + \frac{\alpha - \frac{\beta}{2}}{\pi\beta x'} \right]$$

$$(5.56) \quad = - \frac{\gamma}{\sqrt{2}} \left[\frac{1 - \frac{2}{\beta}}{2 \sinh(\frac{b}{4})} + \frac{\alpha - \frac{\beta}{2}}{\beta} \right].$$

We will compare this to the JT supergravity answer in a moment.

For the two β -ensemble cases, the genus one-half resolvent for Q is (5.15). We can then use $R_H(-z^2) = \frac{1}{2iz}(R_Q(iz) - R_Q(-iz))$ to find the resolvent for H :⁵⁹

$$(5.57) \quad R_{\frac{1}{2}M}(-z^2) = \frac{1 - \frac{2}{\beta}}{z} \int_{-\infty}^\infty \frac{dx'_Q}{x'^2_Q + z^2} \tanh(2\pi x'_Q).$$

⁵⁹One has to take care with $\sqrt{\sigma(x)}$ and $\sqrt{\sigma(x')}$. In the limit where both endpoints of the cut go to infinity, the ratio of these factors is one if x, x' are in the same half-plane, and minus one if they are in opposite half-planes.

After changing variables to $x' = x_Q^2$, we recognize this as two times the expression (5.50) with $\alpha = \frac{\beta}{2}$. So, by comparing to (5.56), the prediction for the volume is

$$(5.58) \quad V_{\frac{1}{2}}(b) = -\sqrt{2}\gamma \frac{1 - \frac{2}{\beta}}{2 \sinh(\frac{b}{4})}.$$

Now, using the values in table 4, and applying (5.56) for the (α, β) -ensemble cases and (5.58) for the β -ensemble cases, one finds that the answer can be written uniformly as

$$(5.59) \quad V_{\frac{1}{2}}(b) = \begin{cases} \frac{\cos \frac{\pi N'}{4}}{\sinh \frac{b}{4}} - \sin \frac{\pi N'}{4} & \widehat{q} = 1 \pmod{4} \\ \frac{\cos \frac{\pi N'}{4}}{\sinh \frac{b}{4}} + \sin \frac{\pi N'}{4} & \widehat{q} = 3 \pmod{4}. \end{cases}$$

This agrees with the computation of the torsion in (3.75), up to the sign that was not fixed in that computation. Note that although a sign was not fixed, it is easy to see from a bulk perspective that the $\widehat{q} = 1$ and $\widehat{q} = 3$ cases should be related by flipping the sign of the second term. This is because the action of \mathbb{T} on the supercharge (and thus the gravitino) has opposite sign in the two cases, which leads to the sign of the η invariant reversing. Since η appears in the combination $\exp(-i\pi\eta N'/2)$, reversing the sign of η is equivalent to reversing the sign of N' , as in (5.59).

Generically, this volume is divergent at $b = 0$ because of the $1/\sinh \frac{b}{4}$ term. This is similar to the divergent $\coth \frac{b}{4}$ term in the purely bosonic case. So for the same reason one finds that the volume of supermoduli space, weighted by the η -invariant factor, is divergent. However, for the special cases with $\beta = 2$ (which correspond to $(\alpha, \beta) = (0, 2)$ and $(2, 2)$), the coefficient of the first term vanishes, and we find $V_{\frac{1}{2}}(b) = \pm 1$. For these two cases, the volume is finite and we can proceed to higher orders.

5.3.3. Higher orders. In fact, the matrix integral prediction for these two cases is that all further correlators vanish in the double-scaled limit. In other words, we claim that for $(\alpha, \beta) = (0, 2)$ or $(2, 2)$, the following is a solution to (4.55):

$$(5.60) \quad \langle R(x) \rangle - \frac{LV'(x)}{2} = e^{S_0} y(x) - \frac{\alpha - 1}{4x}$$

$$(5.61) \quad \langle R(x_1)R(x_2) \rangle_c = \frac{1}{4\sqrt{-x_1}\sqrt{-x_2}(\sqrt{-x_1} + \sqrt{-x_2})^2}$$

$$(5.62) \quad \langle R(I) \rangle_c = 0 \quad |I| \geq 3.$$

Here, we are assuming that the matrix integral is a double-scaled one-cut model, with a single branch point at the origin where $y(x)$ has a $1/\sqrt{-x}$ singularity. We are using a notation appropriate for a double-scaled limit, with both L and e^{S_0} appearing; for the purposes of power-counting, e^{S_0} should be considered of order L .

We can summarize (5.60), (5.61), (5.62) by saying that only terms of order L and order one appear; there are no terms proportional to negative powers of L . To check that we have a solution to the loop equations, one can substitute into (4.55) and consider a $1/L$ expansion. The terms at order L^2 in (4.55) determine $y(x)$ as a function of the potential (or vice versa). The terms at order L determine the order-one pieces of $\langle R(x) \rangle$ and $\langle R(x_1)R(x_2) \rangle_c$. Since (5.60), (5.61), (5.62) do not contain any negative powers of L , all that remains is to check that (4.55) is satisfied at order one. This amounts to checking that the loop equations in the form

$$(5.63) \quad 2xy(x)R_g(x, I) + xF_g(x, I) \sim 0$$

are satisfied for the three cases $(g, |I|) \in \{(1, 0), (\frac{1}{2}, 1), (0, 2)\}$. We will now check each case.

For $(g, |I|) = (1, 0)$, the $F_g(I)$ quantity is

$$(5.64) \quad F_1(x) = \frac{\alpha-1}{2x}R_{\frac{1}{2}}(x) + R_0(x, x) + R_{\frac{1}{2}}(x)^2.$$

For $\alpha \in \{0, 2\}$, this vanishes exactly after using $R_{\frac{1}{2}}(x) = -(\alpha - 1)/(4x)$ and also using the formula for $R_0(x, x_1)$ in (5.61). This means that (5.63) is indeed satisfied with $R_1(x) = 0$. Next, for the case $(g, |I|) = (\frac{1}{2}, 1)$, we have

$$(5.65) \quad F_{\frac{1}{2}}(x, x_1) = \frac{\alpha-1}{2x}R_0(x, x_1) + 2 \left(R_0(x, x_1) + \frac{1}{2} \frac{1}{(x-x_1)^2} \right) R_{\frac{1}{2}}(x)$$

$$(5.66) \quad = -\frac{\alpha-1}{4x(x-x_1)^2}.$$

This expression is nonzero, but $xF_g(x, I)$ is analytic in x in a neighborhood of the cut so (5.63) is satisfied. The final case to consider is $(g, |I|) = (0, 1)$, which determines $R_0(x_1, x_2, x_3)$. This always vanishes in the double-scaled (α, β) -ensembles, as discussed in section 4.2.3.

This matrix integral result, together with the correspondence with JT supergravity, predicts that certain super moduli space volumes are zero. Specifically, the volumes where we sum over pin^- structures, weighting by $\exp(i\pi\eta)$ or by $\exp(-i\pi\eta)$. The first nontrivial case is at genus one with one

boundary, where there is apparently a cancellation between the disk with a handle glued in, and the disk with two crosscaps glued in.

In the matrix integral, we expect that although there are no perturbative corrections in powers of e^{-S_0} , there will be nonperturbative effects; see appendix E.

5.4. Ramond boundaries in the super case

So far, in our discussion of super JT gravity, we have focused on the case that all boundaries have NS spin structure. From the perspective of random matrix theory, any correlator with Ramond spin structures should vanish. For odd N , this is because the Ramond path integral vanishes, as in eqn. (2.8) (or is a constant, still with vanishing correlators, in the ensembles with ν), and for even N it is because for the models we have discussed, the fermionic and bosonic states have precisely the same energies, so $\text{Tr} (-1)^F e^{-\beta H}$ vanishes.

One would like to use this to learn something about the volumes of moduli spaces of super Riemann surfaces with R boundaries.⁶⁰ However, in super JT gravity, this vanishing can actually be explained in a simpler way, just from the fact that the super-Schwarzian path integral gives zero in the Ramond sector. On the disk this is because no R spin structure is possible. For the trumpet, it is because the Schwarzian path integral vanishes (see eqn. (C.11)).

To get around this problem, one can consider instead the quantity $\text{Tr} (Q(-1)^F e^{-\beta H})$, where Q is the supercharge. Now the trumpet super-Schwarzian path integral is nonzero: to leading order for small fluctuations about a classical solution, the supercharge Q of the super-Schwarzian theory is proportional to $S\eta$, where S is the classical value of the bosonic Schwarzian derivative. S is nonzero for the “trumpet” solution, see appendix C, so an insertion of $S\eta$ will soak up the zero mode of η . Since the trumpet is nonzero, if we can show that $\text{Tr} (Q(-1)^F e^{-\beta H})$ has to vanish in a given matrix ensemble, this will imply that in the corresponding bulk theory, volumes of super moduli space with Ramond boundaries must vanish.

In fact, for the matrix ensembles with even N , it is easy to show that $\text{Tr} (Q(-1)^F e^{-\beta H})$ vanishes identically. The operator inside the trace takes bosonic states to fermionic ones and vice versa, so it has no diagonal components and zero trace. For odd N , the situation is more complicated: in this case by “ $\text{Tr} (Q(-1)^F e^{-\beta H})$ ” we really mean a path integral with periodic boundary conditions and an insertion of Q . Such a path integral can

⁶⁰For some anomalies involving these objects, see appendix A.5.

be written in operator language in the Hilbert space of dimension $2^{(N-1)/2}$ where we defined the odd N SYK model. One can show that it reduces to $\pm\sqrt{2}\text{Tr} Qe^{-\beta H}$, with an ambiguous overall sign. For the cases where Q is drawn from an Altland-Zirnbauer ensemble, this vanishes identically because the eigenvalues of Q come in pairs $\omega, -\omega$. For the case where Q is drawn from a GUE ensemble (odd N , no time-reversal), all “stable” correlators vanish, because we argued in section 5.2.1 that all correlators of resolvents of Q beyond $R_0(x)$ and $R_0(x_1, x_2)$ are zero. This argument extends to correlators of any single-trace quantities (including $\text{Tr} Qe^{-\beta H}$), since they can be written in terms of resolvents. This leaves the two cases in which Q is drawn from a GOE-like or GSE-like ensemble. For these cases, all genus-zero contributions will vanish, since GOE and GSE are equivalent to GUE in genus zero. But it seems that in general the correlators with Ramond boundaries are nonzero.

Our conclusion is that correspondence with the matrix integral predicts that volumes of super moduli space with Ramond boundaries must vanish for all of the bulk theories considered, with the exception of unorientable surfaces in the theories with bulk TFT equal to 1 or $\exp(-4i\pi\eta/2)$. For these cases, the volumes will be infinite due to the crosscap divergence, so it seems that there are no cases with finite and nonzero volumes with Ramond boundaries.

5.5. Unbroken supersymmetry and ramond punctures

Several of the ensembles that we have discussed can be generalized to depend on another parameter. This happens if the random matrix C is a bifundamental of a symmetry group $U(L) \times U(L)$, $O(L) \times O(L)$, or $Sp(L) \times Sp(L)$. As discussed in section 2.6.2, these ensembles have a generalization in which C is a bifundamental of $U(L + \nu) \times U(L)$, $O(L + \nu) \times O(L)$, or $Sp(L + \nu) \times Sp(L)$ for some integer ν (an even integer in the Sp case). This corresponds in the ansatz of eqn. (2.63) to a theory with $L + \nu$ bosonic states and L fermionic states. Hence the value of the supersymmetric index is ν , and in fact

$$(5.67) \quad \text{Tr} (-1)^F \exp(-\beta H) = \nu,$$

independent of β . For $\nu \neq 0$, supersymmetry is unbroken; for generic C in this ensemble, there are $|\nu|$ supersymmetric states, of precisely zero energy.

Let us first consider a bifundamental of $U(L + \nu) \times U(L)$, corresponding to a theory with no T symmetry and with even N . By slightly generalizing

the derivation that was given at the end of section 2.1.2, one can show that the eigenvalue measure for such an ensemble takes the Altland-Zirnbauer form (2.3) with

$$(5.68) \quad (\alpha, \beta) = (1 + 2|\nu|, 2).$$

How can JT supergravity be modified to accommodate this and to give a model in which supersymmetry is unbroken? We claim that the parameter ν can be incorporated in JT supergravity by including Ramond punctures.⁶¹ Ideally, we would include Ramond punctures by adding some sort of spin field to the action of JT supergravity. As is usual in string theory, it is difficult to find a good framework for doing that, but there is no problem to incorporate Ramond punctures perturbatively. The recipe is to sum, in every computation, over the number of Ramond punctures that are inserted, weighting each such puncture with a factor of ν . Adding a puncture decreases the Euler characteristic by one, so the total factor associated to such an insertion is

$$(5.69) \quad f = \nu e^{-S_0}.$$

The factor of e^{-S_0} means that a Ramond puncture has the same effect in the genus expansion as a crosscap. In each order of the topological expansion, there is a maximum possible number of Ramond punctures.

Let us first see how this prescription reproduces eqn. (5.67). The lowest order contribution to $\text{Tr}(-1)^F \exp(-\beta H)$ comes from a disc with a single Ramond puncture. The Schwarzian path integral in this situation is equal simply to 1 (eqn. (C.12); note that the factor e^{-S_0} in the Ramond weight eqn. (5.69) was included in writing this formula), and the Euler characteristic of the punctured disk is zero, so the contribution to $\text{Tr}(-1)^F \exp(-\beta H)$ is simply ν , which is the expected result. In the matrix ensemble, the result $\text{Tr}(-1)^F \exp(-\beta H) = \nu$ is an exact formula, with no corrections and no fluctuations in the presence of other operator insertions. The same is true in JT supergravity because any other topology contributing to $\text{Tr}(-1)^F \exp(-\beta H)$ can be constructed by gluing a trumpet, with Ramond spin structure, onto something else. But the Schwarzian path integral of the trumpet vanishes (eqn. (C.11)).

⁶¹A Ramond puncture is not simply the small b limit of a Ramond hole of circumference b , the objects considered in section 5.4. Roughly, a Ramond puncture is obtained by taking $b \rightarrow 0$ and discarding a fermionic mode. See appendix A.3.

The NS sector provides a more elaborate test. Let us first express the partition function $Z_{\text{SJT}}(\beta)$ of SJT supergravity in terms of the corresponding partition function $Z_{\text{M}}(\beta)$ of the Altland-Zirnbauer ensemble with $(\alpha, \beta) = (1 + 2|\nu|, 2)$:

$$(5.70) \quad Z_{\text{SJT}}(\beta) = |\nu| + 2Z_{\text{M}}(\beta).$$

The additive term $|\nu|$ is the contribution of the zero-energy ground states, and as usual the factor of 2 reflects the degeneracy of the energy levels. Equivalently, the SJT and matrix model resolvents are related by

$$(5.71) \quad R_{\text{SJT}}(x) = \frac{|\nu|}{x} + 2R_{\text{M}}(x).$$

A similar formula holds for multi-resolvents.

$R_{\text{SJT}}(x)$ is supposed to be computed by summing over two-manifolds with a single boundary that has NS spin structure. The number of Ramond punctures in such a situation will always be even. Hence only even powers of ν will appear, and in particular, there is no contribution linear in $|\nu|$ to $R_{\text{SJT}}(x)$. At first sight, this appears to contradict eqn. (5.71). But actually, the linear term cancels, as one can learn by evaluating $R_{\text{M}}(x)$. From eqn. (4.68) or (4.69), one sees that shifting α by $2|\nu|$ will shift $R_{\frac{1}{2},\text{M}}$ by $-|\nu|/2x$. Thus

$$(5.72) \quad \widehat{R}_{\frac{1}{2},\text{M}}(x) = -\frac{|\nu|}{2x},$$

where \widehat{R} will denote the lowest order $|\nu|$ -dependent contribution. When this is inserted in eqn. (5.71), the term linear in $|\nu|$ cancels and a contradiction with SJT supergravity is avoided.

In SJT supergravity, the lowest order ν -dependent contribution to the resolvent comes from a disc with two Ramond punctures, with $\chi = -1$. Such a disc can be constructed by gluing a trumpet to a sphere with one hole and two Ramond punctures. The moduli space of super Riemann surface structures on such a sphere is a point, of volume 1. So the contribution of a disc with two punctures to $Z_{\text{SJT}}(\beta)$ is just

$$(5.73) \quad e^{-S_0} \frac{\nu^2}{2} \int_0^\infty b db Z_{\text{SJT}}^T(\beta, b) = \nu^2 e^{-S_0} \sqrt{\frac{\beta}{2\pi}}.$$

We included a factor of ν for each Ramond puncture, and a factor of $1/2$ because a π rotation of the circle on which gluing occurs will exchange the

two Ramond punctures. We also used (C.10) for Z_{SJT}^T . Integrating over β to convert this to a resolvent, we get⁶²

$$(5.74) \quad \widehat{R}_{1,\text{SJT}}(x) = -\frac{\nu^2}{2\sqrt{2}(-x)^{3/2}}.$$

On the matrix model side, the ν^2 term in F_1 is, from eqn. (4.66),

$$(5.75) \quad \widehat{F}_1 = \frac{|\nu|}{x} \widehat{R}_{\frac{1}{2},\text{M}}(x) + \widehat{R}_{\frac{1}{2},\text{M}}(x)^2 = -\frac{\nu^2}{4x^2}.$$

Inserting this in the dispersion relation (4.67) and using

$$y_M = -\cos(2\pi\sqrt{-x})/\sqrt{-2x},$$

we get $\widehat{R}_{1,\text{M}} = -\nu^2/4\sqrt{2}(-x)^{3/2}$. When this is inserted in eqn. (5.71), the result is in accord with the supergravity result (5.74).

This discussion applies to the matrix ensemble appropriate for the theory in which spin structures are weighted with a factor $(-1)^\zeta$. For the theory without $(-1)^\zeta$, one uses a GUE ensemble, and there is no analog of the parameter ν . In appendix A.5, we explain from a bulk point of view why Ramond punctures are only possible in the theory with $(-1)^\zeta$.

Now let us consider the ensembles with symmetry group $O(L + \nu) \times O(L)$ or $\text{Sp}(L + \nu) \times \text{Sp}(L)$. These are appropriate to a \mathbb{T} -invariant theory with N congruent to 0 or 4 mod 8. A similar derivation to the above still applies, with some modifications. First of all, a bifundamental of $O(L + \nu) \times O(L)$ or $\text{Sp}(L + \nu) \times \text{Sp}(L)$ corresponds, respectively, to an Altland-Zirnbauer ensemble with $(\alpha, \beta) = (|\nu|, 1)$ or $(3 + 2|\nu|, 4)$. In each case, the ν -dependent shift in α is $\alpha \rightarrow \alpha + 2\beta|\nu|/\gamma$. Weighting each Ramond puncture by ν , the Ramond sector is as before: a disc with one Ramond puncture makes the desired contribution ν to $\text{Tr}(-1)^F$.

In the NS sector, we now have to use

$$(5.76) \quad R_{\text{SJT}}(x) = \frac{|\nu|}{x} + \gamma R_{\text{M}}(x).$$

A similar calculation to the previous one shows that $\widehat{R}_{\frac{1}{2},\text{M}} = -|\nu|/\gamma x$, as a result of which $R_{\frac{1}{2},\text{SJT}}$ does not depend on ν . This is as expected, since in the NS sector it is not possible to have a single Ramond puncture. For the

⁶²There is no factor e^{-S_0} here merely because the coefficients $R_{g,\text{SJT}}(x)$ in the genus expansion are defined with powers of e^{S_0} removed.

same reason, we expect there to be no contribution linear in $|\nu|$ in $R_{1,\text{SJT}}$. This depends on a cancellation between various terms in $F_{1,\text{M}}$ (eqn. (4.66)) and the identity $(1 - 2/\beta) - (\alpha - 1)/\beta = 0$, which holds at $(\alpha, \beta) = (0, 1)$ or $(3, 4)$. In order ν^2 , we get $F_{1,\text{M}} = -\nu^2/\gamma^2 x^2$. This differs from eqn. (5.75) by a factor of $(2/\gamma)^2$. When we insert $F_{1,\text{M}}$ in the dispersion relation to compute $\widehat{R}_{1,\text{M}}$, we get a factor $\gamma/2$ relative to the previous calculation because for these ensembles, there is an extra factor $2/\gamma$ in y_{M} . Then in $\widehat{R}_{1,\text{SJT}}$, there is another factor of $\gamma/2$ relative to the previous calculation because the 2 on the right hand side of eqn. (5.70) is replaced by γ in eqn. (5.76). The upshot is that $\widehat{R}_{1,\text{SJT}}$ is the same as it was for the $U(L + \nu) \times U(L)$ ensemble. The corresponding SJT computation is also the same as before (since it is not affected by time-reversal symmetry), so again we get agreement between random matrix theory and supergravity.

The analysis in appendix A.5 shows that in the bulk, Ramond punctures are only possible for even N . We expect that a more precise analysis of the anomaly that was involved will show in the unorientable case that N must be congruent to 0 or 4 mod 8, not to 2 or 6.

Acknowledgements

DS was supported in part by Simons Foundation grant 385600. Research of EW is partially supported by NSF Grant PHY-1606531. We are grateful to R. Dijkgraaf, N. Seiberg and S. Shenker for discussions. EW thanks R. Penner for discussions about hyperbolic geometry and super Riemann surfaces and T. Voronov for pointing out some references on supergeometry. We also thank P. Norbury for discussions of his work.

Appendix A. Volumes of symplectic supermanifolds

A.1. Preliminaries

Our goal in this appendix is to obtain a formula in purely bosonic terms for the volume of a symplectic supermanifold – such as the moduli space of super Riemann surfaces. This will make clear some basic properties of those volumes, which we can compare to random matrix theory. It will also enable us to make contact with formulas of Norbury [56], who studied the spectral curve $y = \frac{1}{z} \cos 2\pi z$ (where $z = \sqrt{-x}$) from a different starting point than that of the present paper.

In the bosonic world, a symplectic manifold is a manifold M of dimension $n = 2d$ endowed with a closed two-form ω that is everywhere nondegenerate.

Nondegeneracy means that the volume form

$$(A.1) \quad \text{vol}_M = \frac{\omega^d}{d!}$$

is everywhere nonzero. The volume of M is $\text{Vol}(M) = \int_M \text{vol}_M$, and is manifestly positive if M is oriented appropriately (nondegeneracy of ω implies that M is orientable). An equivalent formula for the volume is

$$(A.2) \quad \text{Vol}(M) = \int_M e^\omega.$$

To justify this, one expands e^ω in powers of ω , and observes that only the term of degree d contributes.

In local coordinates u^1, \dots, u^n , the symplectic form is explicitly $\omega = \frac{1}{2} \sum_{i,j} w_{ij}(u^K) du^i du^j$, where the coefficient matrix w_{ij} is antisymmetric. An alternative formula for the volume form is

$$(A.3) \quad \text{vol}_M = \text{Pf}(w) du^1 du^2 \cdots du^n,$$

where Pf is the Pfaffian.

Now consider a smooth supermanifold \widehat{M} of dimension $n|m$, where $n = 2d$ is even. Suppose that \widehat{M} can be described by local bosonic coordinates u^i , $i = 1, \dots, n$ and local fermionic coordinates θ^a , $a = 1, \dots, m$. The one-forms du^i are treated as fermionic variables and the one-forms $d\theta^a$ are bosonic.⁶³ A two-form $\widehat{\omega}$ is homogeneous and quadratic in these variables. In local coordinates

$$(A.4) \quad \widehat{\omega} = \frac{1}{2} \sum_{i,j} w_{ij} du^i du^j + \sum_{i,a} w_{ia} du^i d\theta^a + \frac{1}{2} \sum_{a,b} w_{ab} d\theta^a d\theta^b,$$

where $w_{ij} = -w_{ji}$ but $w_{ab} = +w_{ba}$. If we write generically X^I for the full set of bosonic and fermionic coordinates $u^i|\theta^a$, and regard w_{ij} , w_{ia} , and w_{ab}

⁶³In supergeometry, one has to decide whether the exterior derivative d commutes or anticommutes with odd constants. The two approaches are equivalent in the sense that results in one can be reformulated in the other; for a precise account, see pp. 62-4 of [94]. However, which approach is more convenient depends on the context. In gauge theory on an ordinary manifold with a supergroup as gauge group – which was the context in our computation of the torsion in section 3.5 – it is more natural to consider the exterior derivative to commute with odd constants. Symplectic geometry on a supermanifold is more straightforward if the exterior derivative is considered to anticommute with odd constants, as we assume here.

as blocks of a supermatrix W , then we can write simply

$$(A.5) \quad \widehat{\omega} = \frac{1}{2} \sum_{I,J} W_{IJ} dX^I dX^J.$$

The closed two-form $\widehat{\omega}$ is said to be nondegenerate if the Berezinian $\text{Ber } W$ is everywhere defined and not nilpotent.⁶⁴ In that case, \widehat{M} is called a symplectic supermanifold with symplectic form $\widehat{\omega}$.

On a supermanifold, we cannot imitate the definition (A.1) of the volume form, because a volume form on a supermanifold is not a differential form (basically because $d\theta$ regarded as a measure transforms oppositely to $d\theta$ regarded as a one-form). There is also no analog of the Pfaffian for a supermatrix. However, we can use the Berezinian, and write

$$(A.6) \quad \text{vol}_{\widehat{M}} = \frac{\sqrt{\text{Ber } W}}{(2\pi)^{m/2}} [du^1 du^2 \cdots du^n | d\theta^1 d\theta^2 \cdots d\theta^m].$$

(The factor of $1/(2\pi)^{m/2}$ will avoid factors of 2π in the final results.) This formula is independent of the local coordinates used – except possibly for an overall sign, coming from the choice of sign of $\sqrt{\text{Ber } W}$ and the ordering of the factors in $[du^1 du^2 \cdots du^n | d\theta^1 d\theta^2 \cdots d\theta^m]$. There can be a topological obstruction to choosing this sign consistently. When this obstruction vanishes, after picking the sign, one defines the volume of \widehat{M} as $\text{Vol}(\widehat{M}) = \int_{\widehat{M}} \text{vol}_{\widehat{M}}$. In contrast to the case of an ordinary symplectic manifold, this volume has no positivity properties. When the sign cannot be defined consistently, one may leave the volume of \widehat{M} undefined or say that it is zero.

We used an arbitrary system of local coordinates in defining $\text{vol}_{\widehat{M}}$. In general, we can cover \widehat{M} by small open sets \mathcal{O}_α , such that in each \mathcal{O}_α , we pick local bosonic and fermionic coordinates $u_\alpha^i, \theta_\alpha^a$. In an intersection of open sets $\mathcal{O}_\alpha \cap \mathcal{O}_\beta$, the local coordinates $u_\alpha^i | \theta_\alpha^a$ are functions of $u_\beta^i | \theta_\beta^a$. One says that the supermanifold \widehat{M} is “split” if the bosonic coordinates u_α^i are

⁶⁴The Berezinian of a supermatrix $W = \left(\begin{array}{c|c} A & B \\ \hline C & D \end{array} \right)$ is infinite, or undefined, if $\det D = 0$, and it is nilpotent if $\det A = 0$. So nondegeneracy of $\widehat{\omega}$ means that the matrices w_{ij} and w_{ab} , which correspond to A and D , both have nonzero determinant. Note that the square root in eqn. (A.6) below does not make sense if $\text{Ber } W$ is nilpotent. Of course, that formula also does not make sense if $\text{Ber } W$ is infinite or undefined.

functions of u_β^j only (and not of θ_β^a) while the θ_α^a are linear in the θ_β^a :

$$(A.7) \quad \begin{aligned} u_\alpha^i &= f_{\alpha\beta}^i(u_\beta^k) \\ \theta_\alpha^a &= \sum_b g_{b\alpha\beta}^a(u_\beta^k)\theta_\beta^b. \end{aligned}$$

In this situation, the u_α^i are local coordinates for a purely bosonic manifold M , and the θ_α^a parametrize the fibers of a vector bundle $V \rightarrow M$. More geometrically, instead of talking about local coordinates, one can say that \widehat{M} is the total space of the vector bundle $V \rightarrow M$, which has purely fermionic fibers.⁶⁵

The fundamental structure theorem of smooth supermanifolds says that every smooth supermanifold can be split.⁶⁶ Thus in analyzing the volume of a smooth symplectic supermanifold \widehat{M} , we can assume that \widehat{M} is the total space of some $V \rightarrow M$. In this situation, the symplectic form $\widehat{\omega}$ of \widehat{M} , when restricted to M (by setting $\theta^a = d\theta^a = 0$) becomes a symplectic form ω on M , and therefore M is an ordinary symplectic manifold. We will see that the volume of \widehat{M} depends only on the cohomology class of the symplectic form ω of M and the topology of the vector bundle $V \rightarrow M$.

Restricted to $\theta^a = 0$, but without setting $d\theta^a = 0$, the symplectic form $\widehat{\omega}$ of \widehat{M} takes the form

$$(A.8) \quad \widehat{\omega} = \frac{1}{2} \left(\sum_{i,j} w_{ij}(u^k) du^i du^j + w_{ab}(u^k) d\theta^a d\theta^b \right).$$

⁶⁵This is sometimes described by saying that \widehat{M} is the total space of $\Pi V \rightarrow M$, where the symbol Π means that the fibers are fermionic. We will just write V rather than ΠV , but we will understand that V is fermionic, that is the fibers of $V \rightarrow M$ are parametrized by fermionic coordinates.

⁶⁶The proof proceeds roughly by starting with any systems of local coordinates $u_\alpha^i|\theta_\alpha^a$, and then trying to improve the coordinates, by adding terms quadratic and higher order in the fermionic variables, so as to put the transition functions in the desired form of eqn. (A.7). The possible obstructions that one encounters are cohomology classes of a smooth manifold with values in a sheaf of smooth functions. Such cohomology vanishes, so \widehat{M} can be split. The splitting is far from unique, but the topology of M , the cohomology class of ω , and the topological type of the vector bundle $V \rightarrow M$ are uniquely determined. (This is most directly proved as follows, without any discussion of coordinates or splitting. M can be defined as the reduced space of \widehat{M} , obtained by reducing \widehat{M} modulo the odd coordinates. As such, M is naturally embedded in \widehat{M} . V is the normal bundle to M in \widehat{M} and ω is the restriction of $\widehat{\omega}$ to M .)

Nondegeneracy of $\widehat{\omega}$ means that the matrices w_{ij} and w_{ab} are both everywhere invertible. In particular, since w_{ab} is symmetric and everywhere nondegenerate, we can view it as a metric tensor on the vector bundle $V \rightarrow M$. This reduces the structure group of V to an orthogonal group $O(m)$ or possibly (if w is not positive- or negative-definite) to an indefinite orthogonal group $O(m_1, m_2)$ where $m_1 + m_2 = m$. We will continue the derivation assuming that w_{ab} is negative-definite, as is the case⁶⁷ for the moduli space of super Riemann surfaces; the generalization does not change much and will be briefly explained at the end. With the nondegenerate metric w_{ab} at hand, we can restrict ourselves to orthonormal systems of local fermionic coordinates, so we can take henceforth $w_{ab} = -\delta_{ab}$.

There is actually no problem with the sign of $\sqrt{\text{Ber } W}$ if w_{ab} is everywhere invertible. However, a sign change of one of the θ^a will reverse the sign of $[du^1 du^2 \cdots du^n | d\theta^1 d\theta^2 \cdots d\theta^m]$ and thus change the sign $\text{vol}_{\widehat{M}}$. So the overall sign of $\text{vol}_{\widehat{M}}$ is reversed if we reverse the orientation of the bundle $V \rightarrow M$. Thus the case that the sign of the volume form can be globally defined is the case that V is orientable, or in other words, that its structure group can be further reduced from $O(m)$ to $SO(m)$. If V is orientable, then the sign of the volume form depends on a choice of orientation of V ; this is an inescapable choice that must be made to define the volume of a symplectic supermanifold.

A.2. Computation

Now we will make use of the projection $\pi : \widehat{M} \rightarrow M$ that forgets the odd coordinates to put the symplectic structure in a convenient form (first described in [101]). The purely bosonic symplectic form

$$\omega = \frac{1}{2} \sum_{i,j} w_{ij}(u^k) du^i du^j$$

⁶⁷The negative-definiteness is visible in Omnibus Theorems B and C of [97]. The symplectic form $\widehat{\omega}$ is described explicitly (in a coordinate system different from that used in the present paper) in Omnibus Theorem B. Its fermionic part is negative-definite, and comparison with Omnibus Theorem C shows that it has been defined so that its bosonic part is the ordinary Weil-Petersson symplectic form ω on the bosonic moduli space. The negative-definite nature of the fermionic part of the symplectic form is also visible in eqn. (C.7) of the present paper, where the bosonic part of the symplectic form related to the super-Schwarzian was described in a way that is consistent with the standard Weil-Petersson symplectic form, and the fermionic part is negative-definite.

can be “pulled back” to a closed, but degenerate, two-form on \widehat{M} that we will call $\pi^*\omega$. Because the fibers of π are contractible (being purely fermionic), the cohomology of M is naturally isomorphic to the cohomology of \widehat{M} . The full symplectic form $\widehat{\omega}$ is therefore equivalent cohomologically to $\pi^*\omega$. Concretely, this means that there is a 1-form λ such that

$$(A.9) \quad \widehat{\omega} = \pi^*\omega + d\lambda.$$

On an ordinary symplectic manifold M , adding an exact form to the symplectic form does not change the volume of M . On a symplectic supermanifold, the same statement holds, as long as the nondegeneracy condition that was needed to define the volume form is maintained. Therefore, as long as the nondegeneracy condition is maintained, the volume of \widehat{M} will not depend on the one-form λ . Because we do need to maintain this nondegeneracy condition, we cannot simply set $\lambda = 0$. But we can choose a convenient λ to simplify the computation of $\text{Vol}(\widehat{M})$.

A convenient λ may be chosen as follows. Pick a gauge connection A on V , with structure group $\text{SO}(m)$. Explicitly, $A = \sum_i A_{iab}(u^k)du^i$, where $A_{iab} = -A_{iba}$. The one-forms $d\theta^a + A_{ib}^a du_i \theta^b$ (where indices are raised and lowered using the metric $w_{ab} = -\delta_{ab}$) are gauge-covariant, so

$$(A.10) \quad \lambda = -\frac{1}{2} \sum_a \theta^a (d\theta^a + A_{ib}^a du_i \theta^b)$$

is gauge-invariant and globally defined. Explicitly, with this choice of λ ,

$$(A.11) \quad \widehat{\omega} = \pi^*\omega - \frac{1}{2} \sum_a d\theta^a d\theta^a + \frac{1}{2} \sum_{a,b} \theta^a \theta^b \partial_i A_{jab} du^i du^j - \theta^a A_{iab} du^i d\theta^b.$$

The volume of a supermanifold with a symplectic structure of this form can be computed rather simply. It is possible to do this directly from eqn. (A.6), but a particularly transparent approach is as follows. First we reconsider an ordinary symplectic manifold M . We view the symplectic form $\omega = \frac{1}{2} w_{ij}(u^k) du^i du^j$ as a function of bosonic variables u^k and fermionic variables du^k . Together these variables parametrize a supermanifold M' which is a fiber bundle over M with fermionic fibers (parametrized by the du^k). Though there is no natural measure on M , there is a natural measure on M' (the integration measures for integrating over the bosonic variable u^k and

the fermionic variable du^k transform oppositely under a change of coordinates, so their product does not depend on a choice of coordinates). Let us call this measure $\mathcal{D}(u^k, du^k)$. The volume of M is then

$$(A.12) \quad \text{Vol}(M) = \int \mathcal{D}(u^k, du^k) \exp\left(\frac{1}{2}w_{ij}(u^k)du^i du^j\right).$$

After thinking through the implications of Berezin integration over the odd variables du^j (the usual definition being $\int \mathcal{D}(du^j) \cdot 1 = 0$, $\int \mathcal{D}(du^j) du^j = 1$ for each j), the reader can hopefully see that eqn. (A.12) is completely equivalent to the definition of the volume in eqn. (A.2). The same approach holds in the case of a supermanifold. If \widehat{M} is a supermanifold with bosonic and fermionic coordinates X^I , then one introduces for each I a new variable dX^I with statistics opposite to those of X^I . Again the X^I and dX^I parametrize a supermanifold \widehat{M}' , and this supermanifold has a natural measure $\mathcal{D}(X^I, dX^I)$. Our previous formula for $\text{Vol}(\widehat{M})$ is equivalent to

$$(A.13) \quad \text{Vol}(\widehat{M}) = \int \frac{\mathcal{D}(X^I, dX^I)}{(2\pi)^m} \exp\left(\frac{1}{2} \sum_{I,J} W_{IJ}(X^K) dX^I dX^J\right).$$

The Gaussian integral over the dX^K generates the factor $\sqrt{\text{Ber } W}$ in the definition of the volume form in eqn. (A.6).

The formula (A.13) is particularly convenient for a symplectic form with the structure in eqn. (A.11), because the Gaussian integral over $d\theta^a$ gives a simple result, after which the integral over the θ^a is also simple. We have explicitly

$$(A.14) \quad \begin{aligned} \text{Vol}(\widehat{M}) = \int_{\widehat{M}'} & \frac{\mathcal{D}(u^i, du^i)\mathcal{D}(\theta^a, d\theta^a)}{(2\pi)^m} \\ & \times \exp\left(\frac{1}{2}\omega_{ij}(u^k)du^i du^j - \frac{1}{2} \sum_a d\theta^a d\theta^a \right. \\ & \left. + \frac{1}{2} \sum_{a,b} \theta^a \theta^b \partial_i A_{j\ ab} du^i du^j - \theta^a A_{i\ ab} du^i d\theta^b\right). \end{aligned}$$

The Gaussian integral over the $d\theta^a$ gives a simple result:

$$(A.15) \quad \text{Vol}(\widehat{M}) = \int \frac{\mathcal{D}(u^i, du^i)\mathcal{D}(\theta^a)}{(2\pi)^{m/2}} \times \exp\left(\frac{1}{2}\omega_{ij}(u^k)du^i du^j + \frac{1}{4}\sum_{a,b}\theta^a\theta^b F_{ij\ ab}du^i du^j\right).$$

Here $F_{ij\ ab} = \partial_i A_{j\ ab} - \partial_j A_{i\ ab} + [A_i, A_j]_{ab}$ are the components of the curvature two-form $F = dA + [A, A]$. The integral over the θ^a gives a gauge-invariant polynomial in F of degree $m/2$. Because of the fermionic nature of the θ^a and the du^i , this polynomial is proportional to what one gets from the product $F_{i_1 i_2 a_1 a_2} F_{i_3 i_4 a_3 a_4} \cdots F_{i_{m-1} i_m a_{m-1} a_m}$ of $m/2$ copies of F by antisymmetrizing in all indices $i_1 i_2 \cdots i_m$ and in all indices $a_1 a_2 \cdots a_m$. Taking into account the explicit $1/(2\pi)^{m/2}$ in eqn. (A.15), this operation builds the Euler class $\chi(V)$, understood as a function of the u^a and du^a that is homogeneous of degree m in the du^a . The only other factor in eqn. (A.15) is $\exp(\frac{1}{2}\omega_{ij}du^i du^j)$. So

$$(A.16) \quad \text{Vol}(\widehat{M}) = \int_{\widehat{M}'} \mathcal{D}(u^i, du^i) \exp\left(\frac{1}{2}\omega_{ij}du^i du^j\right) \cdot \chi(V).$$

A more conventional way to write the same formula is

$$(A.17) \quad \text{Vol}(\widehat{M}) = \int_M \exp(\omega) \cdot \chi(V).$$

Thus we have achieved our goal of expressing the volume of a symplectic supermanifold \widehat{M} purely in terms of bosonic geometry. The result can be compared to explicit computations in examples, such as those in [102].

A.3. Application

In the main example of the present paper, \widehat{M} is the moduli space of super Riemann surfaces, and M is the corresponding reduced space, which is the moduli space of Riemann surfaces with spin structure. In studying this example, it is useful to be able to give V a complex structure. For example, in the case of a Riemann surface Σ without boundary, M is a complex manifold (the moduli space that parametrizes a Riemann surface Σ with a choice of a square root $K^{1/2}$ of its canonical bundle K). The fiber of the bundle V is $H^1(\Sigma, K^{-1/2})$ (which parametrizes the fermionic directions in \widehat{M}). So in

particular in this case, V is a holomorphic vector bundle, not just a complex vector bundle. If Σ has boundaries, then M is no longer a complex manifold so it does not make sense to say that V is holomorphic. But for boundaries of NS type, one can still show that V can be given the structure of a complex vector bundle. The explanation of this is somewhat subtle and will be given below, along with an explanation that $\chi(V) = 0$ if there are boundaries of Ramond type. Thus in our main example, we can always assume that V is a complex vector bundle. A vector bundle with a complex structure acquires a natural orientation, so the complex structure of V gives us a way to define the sign of $\chi(V)$. In fact, the class that is naturally called $\chi(V)$ if V is viewed as a real vector bundle of some rank m is equivalently the top Chern class $c_{m/2}(V)$ if V is viewed as a complex vector bundle of rank $m/2$.

This gives a succinct way to orient the bundles V for all Σ in a uniform fashion, and moreover in a way that behaves well under gluing. But it is important to point out that a choice is involved. If J is a complex structure on V , then $-J$ is an equally good complex structure. For example, if one choice identifies V with $H^1(\Sigma, K^{-1/2})$, the other choice will identify V with the dual complex vector bundle, which is $H^0(\Sigma, K^{3/2})$. Reversing the complex structure of V will multiply the orientation of V by $(-1)^{m/2}$, where m is the rank of V (the odd dimension of the symplectic supermanifold \widehat{M}). In our problem of genus g super Riemann surfaces with s boundaries, $m = 4g - 4 + 2s$ so $(-1)^{m/2} = (-1)^s$. Purely from the point of view of symplectic supergeometry, two formalisms differing by a factor of $(-1)^s$ are equally natural. In this paper, we resolve this ambiguity to make the disc partition function positive (as one expects from reflection positivity of supergravity). As a disc has $s = 1$, the sign of the disc amplitude does resolve the ambiguity. The sign of the measure of a three-holed sphere is also sensitive to the same choice. We only fully fix this sign in appendix D.7, where we fix the meaning of the symbol $[d\xi d\psi]$ by using $\int [d\xi d\psi] \xi\psi = +1$ in the derivation of eqn. (D.42). The choice amounted to a choice of orientation of the bundle V , for the case of a three-holed sphere. Once this choice is made, the measure for all other cases is determined by gluing (eqn. (3.100)). Our choices for the disc and the three-holed sphere are compatible; this is visible in the fact that the results deduced in appendix D, both for the recursion relation and for the input $V_1(b)$ of the recursion relation, agree with the results of section 5.2.

As explained in footnote 48 of section 3.5.4, we have defined the metric on the $\mathfrak{osp}(1|2)$ Lie algebra so that the symplectic form $\widehat{\omega}$ on \widehat{M} restricts on M to the standard Weil-Petersson symplectic form ω . On the other hand, the “triviality” of the torsion on an orientable two-manifold means that the measure τ on \widehat{M} that comes from the torsion is $\tau = \sqrt{\text{Ber } \widehat{\omega}}$. To get the

natural formula (A.17) with no factors of 2π , we started in eqn. (A.6) not with $\sqrt{\text{Ber } \widehat{\omega}}$ but with $\sqrt{\text{Ber } \widehat{\omega}}/(2\pi)^{m/2}$, where m is the odd dimension of \widehat{M} . In our application to super Riemann surfaces, $m = -2\chi(\Sigma)$. So a measure μ that leads to the natural normalization (A.17) of the volumes is actually not τ but

$$(A.18) \quad \mu = (2\pi)^\chi \tau.$$

The formula (A.17) has a number of interesting corollaries. If m is odd, then $\chi(V)$ vanishes trivially at least modulo torsion (as it is supposed to be a polynomial of degree $m/2$ in F), so a symplectic supermanifold with an odd fermionic dimension has zero volume. If the rank of the vector bundle V exceeds the dimension of M , then the class $\chi(V)$ vanishes for dimensional reasons, so a supermanifold of dimension $n|m$ with $m > n$ has zero volume. In general, as $\chi(V)$ is a class of degree m , the volume $\text{Vol}(\widehat{M})$ is proportional to $\omega^{(n-m)/2}$.

These statements have interesting implications for our problem of hyperbolic Riemann surfaces with boundary, in which \widehat{M} has dimension $6g - 6 + 2s|4g - 4 + 2s$. In particular, if $g = 0$, this is of the form $n|m$ with $n < m$, so the volumes vanish. Now let us discuss what happens for $g > 0$. The topological type of the vector bundle V and the integral cohomology class $\chi(V)$ are determined by discrete data, so they are independent of the b_i . Hence the dependence of $\text{Vol}(\widehat{M})$ on the b_i comes entirely from the dependence of ω on the b_i . Mirzakhani showed that the cohomology class of ω is a linear function of the variables b_i^2 . (Her analysis was phrased for bosonic Riemann surfaces, but is not significantly affected by incorporating a spin structure.) With \widehat{M} having dimension $n|m = 6g - 6 + 2s|4g - 4 + 2s$, its volume is a polynomial in ω of degree $(n - m)/2 = g - 1$. So $\text{Vol}(\widehat{M})$ will be a polynomial in the b_i^2 of that degree. We do not have a general proof of this behavior from the matrix model, but the examples in eqn. (5.40) do have the claimed property.⁶⁸

Now let us explain why V can be given a complex structure in the case of NS boundaries, and why $\chi(V)$ vanishes when there are Ramond boundaries. A hyperbolic conjugacy class in $\text{OSp}(1|2)$ with spin structure of NS type

⁶⁸This property has been proved by Norbury in unpublished work, using the recursion relation (5.42), which he discovered independently of the present paper, by extending the topological recursion constructed in [56].

contains an element

$$(A.19) \quad U_b = \left(\begin{array}{cc|c} e^{b/2} & 1 & 0 \\ 0 & e^{-b/2} & 0 \\ \hline 0 & 0 & -1 \end{array} \right).$$

where $b > 0$ is a length parameter. For any positive b , the matrix element of U_b above the diagonal could be removed by conjugation. The reason for including it is that this ensures that the hyperbolic conjugacy class that contains U_b has a smooth limit for $b \rightarrow 0$. In that limit, U_b goes over to

$$(A.20) \quad U_0 = \left(\begin{array}{cc|c} 1 & 1 & 0 \\ 0 & 1 & 0 \\ \hline 0 & 0 & -1 \end{array} \right),$$

which is a parabolic element of $\mathrm{OSp}(1|2)$. The conjugacy class of $\mathrm{OSp}(1|2)$ containing U_0 has the same dimension, namely $2|2$, as the hyperbolic conjugacy class that contains U_b , and the conjugacy class containing U_b goes over smoothly for $b \rightarrow 0$ to the conjugacy class containing U_0 . U_0 is the monodromy of a flat $\mathrm{OSp}(1|2)$ connection around an NS puncture (not hole). So in the NS case, a hole of length b goes over smoothly to a puncture for $b \rightarrow 0$. In the limit that the holes are replaced by punctures, the moduli space of super Riemann surfaces becomes a complex manifold, and therefore V gets a complex structure. For questions (like computing the integral cohomology class $\chi(V)$) that are independent of continuously variable parameters such as the lengths, we can set the length parameters of NS boundaries to 0. Therefore, in the case of NS boundaries, V can be given a complex structure, as was claimed earlier.

For holes of Ramond type, a corresponding argument shows instead that the volumes vanish, as we found in another way in section 5.4. A hyperbolic conjugacy class with Ramond spin structure contains an element

$$(A.21) \quad \tilde{U}_b = \left(\begin{array}{cc|c} e^{b/2} & 1 & 0 \\ 0 & e^{-b/2} & 0 \\ \hline 0 & 0 & 1 \end{array} \right).$$

for some $b > 0$. Naively setting $b = 0$, we get a parabolic element of $\mathrm{OSp}(1|2)$:

$$(A.22) \quad \tilde{U}_0 = \left(\begin{array}{cc|c} 1 & 1 & 0 \\ 0 & 1 & 0 \\ \hline 0 & 0 & 1 \end{array} \right),$$

\tilde{U}_0 is the monodromy of a flat $\mathrm{OSp}(1|2)$ connection around a Ramond puncture (as opposed to a Ramond boundary). The limit from \tilde{U}_b to \tilde{U}_0 is not smooth. The orbit of \tilde{U}_b in $\mathrm{OSp}(1|2)$ is of dimension $2|2$, but the orbit of \tilde{U}_0 is of dimension $2|1$, because \tilde{U}_0 commutes with a certain linear combination of the odd generators of $\mathrm{OSp}(1|2)$. Hence, a moduli space of super Riemann surfaces with a Ramond boundary of length b does not approach a corresponding moduli space with a Ramond puncture in the limit $b \rightarrow 0$. If one sets $b = 0$, one of the odd moduli is lost. Thus if V parametrizes the odd moduli of a super Riemann surface with a specified Ramond boundary of length b (along with other boundaries), and V' parametrizes the odd moduli if the boundary in question is replaced by a Ramond puncture, then the rank of V' is 1 less than the rank of V . In the limit $b \rightarrow 0$, one can define a “forgetful” map $V \rightarrow V'$ in which one identifies two super Riemann surfaces that are equivalent at $b = 0$. If V'' is the kernel of this forgetful map, then V'' is a real vector bundle of rank 1, and there is an exact sequence $0 \rightarrow V'' \rightarrow V \rightarrow V' \rightarrow 0$. This implies that $\chi(V) = \chi(V'')\chi(V')$. But $\chi(V'') = \chi(V') = 0$ (as rational cohomology classes) since V' and V'' are of odd rank, so $\chi(V) = 0$. Let us note that this argument also shows that correlators with Ramond punctures (as opposed to holes) are essentially new quantities. They are analyzed in section 5.5.

The bundle V can be given a natural complex structure in the case of Ramond punctures, but not in the case of Ramond boundaries.

Two generalizations are worth mentioning. First, we assumed that w_{ab} was negative-definite. More generally, we can decompose V as the direct sum $V_1 \oplus V_2$ of subbundles on which w_{ab} is positive-definite or negative-definite, respectively. If V_1 and V_2 have ranks m_1 and m_2 , respectively, one can make a similar derivation, leading to the result that $\chi(V)$ is replaced by $(-1)^{m_1/2}\chi(V_1)\chi(V_2) = (-1)^{m_1/2}\chi(V)$. Second, we have implicitly assumed that the smooth supermanifold \widehat{M} is real, meaning that the fermions carry a real structure and the vector bundle V is real. A generalization that has been called a cs supermanifold is important in superstring theory [94, 103]. Here the fermions do not carry any real structure. However, in the symplectic case, the generalization from real supermanifolds to cs supermanifolds does not add much, in the following sense. *A priori*, if \widehat{M} is a cs supermanifold, the structure group of $V \rightarrow M$ is the complex linear group $\mathrm{GL}(m, \mathbb{C})$. A general $\mathrm{GL}(m, \mathbb{C})$ bundle cannot be reduced to a bundle with real structure group $\mathrm{GL}(m, \mathbb{R})$, so in general a cs supermanifold cannot be given a real structure. But in the case of a symplectic cs supermanifold, the metric tensor w_{ab} reduces the structure group from $\mathrm{GL}(m, \mathbb{C})$ to the complex orthogonal group

$O(m, \mathbb{C})$. A further reduction to a maximal compact subgroup is always possible. A maximal compact subgroup of $O(m, \mathbb{C})$ is the ordinary compact orthogonal group $O(m)$. The m -dimensional representation of this group is real, showing that if \widehat{M} is a symplectic cs supermanifold, it actually can always be given a real structure.

A.4. Comparison to the results of Norbury

In [56], Norbury studied the spectral curve $y = \frac{1}{z} \cos(2\pi z)$ (with $z = \sqrt{-x}$) from a different starting point than the one of the present paper. Seeking to generalize the results of Eynard and Orantin for the spectral curve $y = \sin(2\pi z)$ (which they had shown to be related to formulas of Mirzakhani for volumes of moduli spaces of Riemann surfaces), Norbury was led to introduce the space M that parametrizes a Riemann surface with a choice of spin structure, along with the vector bundle that we have called V . He found that the analog for the “cosine” curve of the usual moduli space volumes for the “sine” curve were the quantities that appear on the right hand side of eqn. (A.17). The sense in which these quantities actually are volumes was unclear, but we have provided an answer: they are the volumes of the supermanifolds \widehat{M} that parametrize super Riemann surfaces. Norbury’s class $\Theta_{g,s}$ is defined as (the sum over spin structures of) the top Chern class of a vector bundle $E_{g,s}$ whose fiber is $H^0(\Sigma, K^{3/2})$ (see Definition 2.1 in [56]).

The analysis in section D.4 makes it clear that to define the volume of a symplectic supermanifold, one needs to pick an orientation of the vector bundle V that parametrizes the fermionic moduli. To compare our results to Norbury’s, it is necessary to compare the two orientation conventions, which in fact do not agree.

The two choices are as follows:

(1) In this paper, after picking a measure on the Lie superalgebra $\mathfrak{osp}(1|2)$, we use the torsion to orient the moduli spaces. Differently put, the torsion has a sign, and this sign is an orientation of the bundle V . Though it involves more machinery than a purely symplectic approach, the torsion has the advantage of making sense on an unorientable two-manifold. It also has nice gluing properties, in the spirit of topological field theory. For a two-manifold Y , let or_Y be the orientation of V determined by the torsion (in other words, or_Y is the orientation of V that matches the sign of the torsion measure τ_Y). If Y is built by gluing two-manifolds Y_1, Y_2 along one or more circles, then the gluing law for the torsion gives

$$(A.23) \quad \text{or}_Y = (-1)^{w_{\text{R}}} \text{or}_{Y_1} \text{or}_{Y_2}.$$

Here w_R is the number of circles with Ramond spin structure on which gluing has occurred. Like the similar factor in eqn. (3.100), this factor comes from the sign of the torsion of a circle (eqn. (3.69)).

(2) For orientable Y , Norbury orients V by using the complex structure in which $V = H^0(Y, K^{3/2})$. As we explained in appendix A.3, a variant of this procedure would be to use the opposite complex structure on V . This would multiply the orientation (and hence the volume) for a genus g surface with s boundary components by a factor $(-1)^s$.

To see that these two choices do not agree, it suffices to consider the example of a genus 1 surface with a single boundary component. In eqn. (5.40), we found that if we sum over spin structures with a factor $(-1)^\zeta$, we get $V_1(b) = -1/8$, while at the end of section 5.2.1, we showed that a similar sum without the factor $(-1)^\zeta$ gives $V_1(b) = 0$. The two results together show that with our orientations, the moduli spaces with even and odd spin structures (and not including a factor $(-1)^\zeta$ in the odd case) have volumes $-1/16$ and $+1/16$, respectively.

On the other hand, Norbury shows that in genus 1 with a single boundary component, and summing over spin structures with no $(-1)^\zeta$, $\int_{\mathcal{M}_{1,1}} c_1(E_{1,1}) = +1/8$ (see Proposition 2.10, and note that $\int_{\mathcal{M}_{1,1}} \psi_1 = 1/24$). Moreover the even and odd spin structures make equal contributions to this integral.⁶⁹ So the integral for either the even or odd spin structure is $+1/16$.

Thus our results agree with Norbury's for the case of an odd spin structure, but disagree for the case of an even spin structure. Accordingly, our results will differ from Norbury's by a factor $(-1)^s(-1)^\zeta$. The factor $(-1)^s$ is inessential, in the sense that it could be removed by using the opposite complex structure on V . But the factor $(-1)^\zeta$ really is essential.

In appendix A.5, we will show in another way why the two approaches must differ by a factor $(-1)^\zeta$.

A.5. Ramond punctures and an anomaly

In the presence of Ramond punctures (rather than boundaries), the bundle V has a complex structure, and this gives a natural way to orient it. In particular, the Ramond punctures behave as identical bosons; one can define the volumes without specifying how they should be ordered. Thus, in approach

⁶⁹This was explained to us by P. Norbury. Note that $\int_{\mathcal{M}_{1,1}} c_1(E_{1,1})$ can be a rational number (rather than an integer) because the moduli space is an orbifold, not a smooth manifold.

(2) to orienting the moduli spaces, the volumes with Ramond punctures are manifestly well-defined, as long as one does not include a factor $(-1)^\zeta$.

However, once Ramond punctures are present, there is no distinction between even and odd spin structures, since the monodromy that occurs when a Ramond puncture loops around a one-cycle in a two-manifold Y can permute the spin structures of Y in an arbitrary fashion.⁷⁰ Hence it is not possible to define the factor $(-1)^\zeta$ in the presence of Ramond punctures, and therefore in approach (2), if Ramond punctures are present, the volumes can only be defined without the factor $(-1)^\zeta$.

If one considers Ramond boundaries rather than Ramond punctures, the situation is opposite. The difference between a Ramond boundary and a Ramond puncture, as explained in appendix A.3, is that replacing a Ramond puncture with a very small Ramond boundary adds one fermionic modulus, which is localized near the boundary. So as Ramond punctures behave as identical bosons when approach (2) is used to orient the moduli spaces, Ramond boundaries behave as identical fermions in that approach. This is an anomaly of sorts. It means that the volume would depend on an ordering of the Ramond boundaries, modulo even permutations. This dependence can be canceled by supplying a factor of $(-1)^\zeta$. To see this, note that another way to associate a fermionic mode to a Ramond boundary would be to introduce a Majorana fermion that propagates on this boundary. This is the anomaly that is canceled by a bulk factor $(-1)^\zeta$, as we discussed in section 2.4.

In approach (1) to orienting the moduli spaces, matters are reversed. The torsion with possible Ramond boundaries was analyzed in section 3.5 and did not depend on any ordering of the boundaries; the Ramond boundaries behaved like identical bosons. So Ramond boundaries make sense in approach (1) to orienting the moduli spaces with no factor $(-1)^\zeta$. Ramond boundaries become anomalous if one tries to include a factor $(-1)^\zeta$ in the sum over spin structures, because this causes the Ramond boundaries to behave as fermions.

What if we replace Ramond boundaries with Ramond punctures? This entails removing a fermionic mode from every Ramond boundary, so if the Ramond boundaries behave as bosons, the Ramond punctures behave as fermions, and vice-versa. This means, in approach (1), that the Ramond

⁷⁰A spin structure in the presence of Ramond punctures is a spin structure on the complement of the punctures that has a certain type of singularity at the punctures. (A fermion that is parallel transported on a small loop around a Ramond puncture has monodromy -1 .) When such a singularity moves around a 1-cycle γ , the monodromy for a fermion path that has an odd intersection number with γ is reversed in sign.

punctures are bosonic in the theory with $(-1)^\zeta$ and fermionic in the theory without it. (It would be possible to show this explicitly by extending the torsion computations of section 3.5 to include Ramond punctures.) Hence in approach (1), in the presence of Ramond punctures, the volumes can be defined in the theory with $(-1)^\zeta$ but not in the theory without it.

In this discussion, we have treated Ramond punctures and Ramond boundaries in parallel. However, in the context of the present paper, there is an important difference. We treat Ramond punctures (when we do include them) as an intrinsic part of the theory; their positions are some of the moduli that parametrize the moduli spaces whose volumes we compute. If the Ramond punctures behave as fermions, the volumes cannot be defined and the theory is inconsistent. But we treat Ramond (or NS) boundaries as external probes of the theory. If they behave as fermions, these external probes have unexpected properties, but the theory does not become inconsistent. We could eliminate this asymmetry by incorporating in the theory D-branes and thus dynamical boundaries. If a dynamical Ramond boundary is a fermion, the theory is anomalous or inconsistent. Hence, which D-branes are possible will depend on whether a factor $(-1)^\zeta$ is present. This should not be a surprise, since the same is true for related theories such as the 0A and 0B theories that are briefly described in appendix F.

Appendix B. Behavior under $N \rightarrow N + 4$

For any of the ten matrix ensembles discussed in section 2.1, after picking a suitable matrix potential, one gets a matrix integral that can be expanded in Feynman diagrams. The diagrams can be conveniently drawn using the “double line” notation of ’t Hooft [77], in which a propagator is drawn as a thin ribbon. The edges of the ribbon represent the indices of the matrix, and join into “index loops.” By gluing the boundary of a disc onto each index loop in such a diagram, one can build a two-manifold. This is the starting point for the relation between random matrices and two-dimensional gravity.

For the most basic case of a hermitian matrix with unitary symmetry – related to GUE statistics – the two-manifolds constructed from Feynman diagrams are orientable. The same is true for one of the Altland-Zirnbauer ensembles: the bifundamental of $U(L) \times U(L)$. It is undoubtedly not a coincidence that this particular ensemble appears to be related to the volumes of moduli spaces of orientable super Riemann surfaces. This fact, and its analog for other cases, suggests that Altland-Zirnbauer ensembles are somehow related to random supergeometries just as the Dyson ensembles are related to random geometries. But it is unclear how to make this precise.

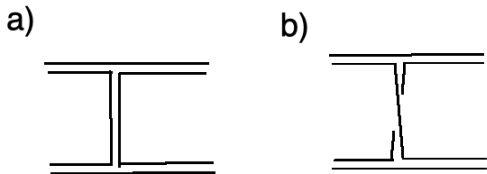


Figure B1: Drawn in a) and b) are portions of two Feynman diagrams that differ only by a relative “twist” in one propagator – the one drawn vertically. When these pictures are embedded in a more complete Feynman diagram, the twist will change the number of index loops and therefore the Euler characteristic of the resulting two-manifold by ± 1 .

The Feynman diagram expansion of a model based on the bifundamental of $U(L) \times U(L)$ was actually studied some time ago [98]; the results were not related to supergeometry in an obvious way. An alternative idea is that discretization of supergeometry might be described by a random ensemble of supermatrices; see [104] for recent work on supermatrices.

For the other eight ensembles, unorientable as well as orientable two-manifolds appear. These eight ensembles all enter, as we have reviewed in section 2.6.2, in the random matrix classification of T -invariant models, with or without supersymmetry. In that context, the eight ensembles are exchanged in pairs under $N \rightarrow N + 4$. (For example, for the case of a supersymmetric matrix model with $N = 1$ or $5 \pmod{8}$, $N \rightarrow N + 4$ exchanges GOE and GSE.) As we discussed in section 2.5.3, from a topological field theory point of view, one expects the exchange $N \rightarrow N + 4$ to multiply the contribution of a two-manifold of Euler characteristic χ by $(-1)^\chi$.

One would hope to find the same behavior in the Feynman diagram expansion of a matrix integral. This appears to be the case, but the literature does not seem to contain a complete proof. For the basic case of comparing GOE-like and GSE-like ensembles, a proof was given in [60]. This proof was surprisingly complicated. The difficulty of the proof comes from the fact that in this case, as $N \rightarrow N + 4$ changes the symmetry group, it changes the vertices as well as the propagators in a Feynman diagram.

We will not summarize this proof, and instead will just point out a pair of Altland-Zirnbauer ensembles that differ by $N \rightarrow N + 4$ and can be compared in a simple way because the symmetry groups are the same. We consider the two models in which the matrix C is a symmetric or antisymmetric second

rank tensor of $G = U(L)$ (corresponding to supersymmetric models with $N = 2, 6 \pmod 8$). The matrix action is $\text{Tr } f(CC^\dagger)$ for a suitable function f . The matrix propagator is, up to a possible constant multiple,

$$(B.1) \quad \langle C_{ij} C^{\dagger kl} \rangle = \frac{1}{2} \left(\delta_i^k \delta_j^l \pm \delta_j^k \delta_i^l \right),$$

with a $+$ or $-$ sign depending on whether C is a symmetric or antisymmetric tensor. The only difference between the Feynman diagram expansions of the two models comes from this sign. Now consider two Feynman diagrams that differ only by which term we take in one of the propagators. In the double line notation, they differ only by whether an extra “twist” is given to the ribbon that corresponds to this propagator (fig. B1). Adding or removing such a twist changes the number of index loops by ± 1 , so it shifts χ by ± 1 . If the tensor C is symmetric, adding or removing a twist does not affect the sign of the diagram, but if C is antisymmetric, it does, because of the minus sign in eqn. (B.1). So the Feynman diagram expansion of the antisymmetric tensor model differs from that of the symmetric tensor model by an extra factor $(-1)^\chi$, as expected.

Appendix C. Super-Schwarzian path integrals for the disk and trumpet

As explained for the bosonic case in [35], the JT path integral reduces to an integral over the moduli space of bordered (super) Riemann surfaces, together with a path integral over the “boundary wiggles” associated to the asymptotically AdS boundary. In this section, we will compute the path integral over the boundary wiggles. There are two cases to consider: the case where the wiggles are at the boundary of the hyperbolic disk, and the case where they are at the “big end” of a hyperbolic trumpet whose other end connects to a bordered Riemann surface, see figure 17.

In the bosonic case, these boundary wiggles are governed by the Schwarzian theory, and the measure that follows from the JT path integral is the natural symplectic measure with respect to which the Schwarzian theory is one-loop exact [35]. In the supersymmetric case, the wiggles (together with a fermionic superpartner) are governed by the super-Schwarzian theory [30] introduced in [81]. In principle, for super JT gravity, one could follow the steps in the bosonic case and determine the associated measure by starting from the formulation of JT supergravity as an $OSp(1|2)$ BF theory. We will assume that the measure that one obtains from this procedure is the natural symplectic one that makes the path integral one-loop exact.

The purely bosonic part of the super-Schwarzian action is the ordinary Schwarzian derivative $\text{Sch}(f(u), u)$, where u is the Euclidean time coordinate of the boundary theory, running from zero to β . The u coordinate measures rescaled proper time along the wiggly boundary:

$$(C.1) \quad ds_{\parallel} = \frac{du}{\epsilon},$$

where ϵ is interpreted as a holographic renormalization parameter, see [5]. The total length of the wiggly boundary is therefore β/ϵ . The detailed shape of the boundary is specified by the function $f(u)$. This gives the Euclidean ‘‘Poincare time’’ of the wiggly boundary as a function of the coordinate u . More precisely, the f coordinate is such that the metric of hyperbolic space is

$$(C.2) \quad ds^2 = \frac{df^2 + dz^2}{z^2}.$$

Specifying $f(u)$ and (C.1) specifies the boundary curve, up to isometries. There are two special cases that will be important for us. First, we have the classical solution corresponding to a circular boundary for the disk, which is $f = \tan(\pi u/\beta)$ up to $SL(2, \mathbb{R})$ transformations (which becomes $OSp(1|2)$ in the super-Schwarzian case). For this solution, $\text{Sch}(f(u), u) = 2\pi^2/\beta^2$. Second, we have the classical solution corresponding to a circular boundary for the trumpet, which is $f = \tanh(\frac{bu}{2\beta})$ up to $U(1)$ transformations, with $\text{Sch}(f(u), u) = -b^2/(2\beta^2)$.

The super-Schwarzian theory [81] adds to this bosonic piece a fermion η that is coupled to f . A useful fact is that, with respect to the symplectic measure, the path integral of the super-Schwarzian theory is one-loop exact [15], so we can determine it by analyzing the one-loop path integral around the relevant classical solution. Let’s imagine that $\{f = f_0(u), \eta = 0\}$ is a classical solution. The equation of motion implies that $\text{Sch}(f_0(u), u)$ is constant, and we will refer to its constant value as S . Then we expand near this solution, by making a small reparametrization $f(u) = f_0(u + \varepsilon(u))$.⁷¹ To quadratic order in ε and the fermion η , the super-Schwarzian action is [81]

$$(C.3) \quad I = -\gamma \int_0^\beta du \left(S - \frac{1}{2} \varepsilon'^2 + S \varepsilon'^2 + 2\eta' \eta'' - S \eta \eta' \right).$$

⁷¹This $\varepsilon(u)$ is unrelated to the holographic renormalization parameter ϵ .

Here γ is a constant with dimensions of length that can be absorbed by rescaling β (this is equivalent to an overall rescaling of the matrix in the dual matrix integral). We will follow the conventions of [35] and choose units such that $\gamma = \frac{1}{2}$. Then β becomes dimensionless.

To evaluate the one-loop path integral, it is convenient to expand in Fourier modes

$$\begin{aligned}
 \varepsilon(u) &= \sum_n e^{-\frac{2\pi}{\beta}inu} \left(\varepsilon_n^{(R)} + i\varepsilon_n^{(I)} \right), \\
 \eta(u) &= \sum_m e^{-\frac{2\pi}{\beta}imu} \left(\eta_m^{(R)} + i\eta_m^{(I)} \right).
 \end{aligned}
 \tag{C.4}$$

We take ε to be periodic and η to be antiperiodic, so $n \in \mathbb{Z}$ and $m \in \mathbb{Z} + \frac{1}{2}$. In order for the fields to be real, the real parts $\varepsilon^{(R)}, \eta^{(R)}$ should be even in n and m , and the imaginary parts $\varepsilon^{(I)}, \eta^{(I)}$ should be odd. In these variables, the action (C.4) is

$$I = -\frac{\beta}{2}S + \sum_{n>0} \frac{\beta}{2} \left(\left(\frac{2\pi n}{\beta} \right)^4 - 2S \left(\frac{2\pi n}{\beta} \right)^2 \right) \left((\varepsilon_n^{(R)})^2 + (\varepsilon_n^{(I)})^2 \right)
 \tag{C.5}$$

$$- \sum_{m>0} 4\beta \left(\left(\frac{2\pi m}{\beta} \right)^3 - \frac{S}{2} \frac{2\pi m}{\beta} \right) \eta_m^{(R)} \eta_m^{(I)}.
 \tag{C.6}$$

In general, this action has zero-modes. For a generic value of S , the only zero mode is the $n = 0$ mode of ε , but for special values of S , there can be more. In the super JT theory, these zero modes are interpreted as pure gauge modes. For example, rotations of the wiggly boundary of the trumpet relative to an arbitrary reference coordinate system, or $\text{OSp}(1|2)$ transformations of the wiggly boundary of the disk relative to an arbitrary reference. We should not integrate over these modes.

An important fact for the one-loop exactness of the super Schwarzian theory is that the quotient of the path integral space by the space of these zero modes is a symplectic supermanifold. The symplectic form is the Kirillov-Kostant-Souriau form for the coadjoint orbits of super Virasoro, and it has the property that with respect to the symplectic form, the Schwarzian action is a Hamiltonian that generates the $U(1)$ symmetry of translations in u . To quadratic order in the fields, the form is the restriction to nonzero

modes of

$$(C.7) \quad \widehat{\omega} = \frac{\alpha}{2} \int_0^\beta du \left[d\varepsilon'(u) d\varepsilon''(u) - 2S d\varepsilon(u) d\varepsilon'(u) - 4d\eta'(u)^2 + 2S d\eta(u)^2 \right].$$

A quick way to justify this is to check that with respect to this form, the action (C.3) generates u -translations. Concretely, one needs to show that dH is proportional to the interior product of the u -derivative vector field with $\widehat{\omega}$. This is computed by replacing one of the factors of $d\varepsilon$ and $d\eta$ by $\varepsilon'(u)$ and $\eta'(u)$.

The constant α that multiplies (C.7) is arbitrary, but it is correlated with a similar arbitrary constant in the normalization of the symplectic form on supermoduli space (in the orientable case). This is due to the fact that both symplectic forms are restrictions of the same underlying symplectic form of BF theory. Changing the overall normalization of this form corresponds to the freedom to adjust the coefficient of the topological term in the JT action. We will follow the convention in [35] and set the constant equal to $\alpha = 1$. The bosonic computation in [35] implies that the corresponding normalization for the symplectic form on super moduli space is such that the bosonic component is $\sum_i db_i d\tau_i$ in terms of the Fenchel-Nielsen (length-twist) coordinates. Setting $\alpha = 1$ and inserting the Fourier expansion, one finds

$$\begin{aligned} \widehat{\omega} = & \sum_{n>0} 2\beta \left(\left(\frac{2\pi n}{\beta} \right)^3 - 2S \frac{2\pi n}{\beta} \right) d\varepsilon_n^{(R)} d\varepsilon_n^{(I)} \\ & - \sum_{m>0} 4\beta \left(\left(\frac{2\pi m}{\beta} \right)^2 - \frac{S}{2} \right) \left((d\eta_m^{(R)})^2 + (d\eta_m^{(I)})^2 \right). \end{aligned}$$

With these formulas, it is straightforward to evaluate the Gaussian integral

$$(C.8) \quad \int \prod_{n>0} \mathcal{D}(\varepsilon_n^{(a)}, d\varepsilon_n^{(a)}) \prod_{m>0} \frac{\mathcal{D}(\eta_m^{(a)}, d\eta_m^{(a)})}{(2\pi)^2} \exp(\widehat{\omega} - I) \\ = e^{\frac{\beta}{2} S} \prod_{n>0} \frac{2\beta}{n} \prod_{m>0} \frac{m}{2\beta}.$$

We are interested in applying this formula to two different cases. The first case is the disk partition function, for which the classical solution has $S = 2\pi^2/\beta^2$. For this case, the action has three bosonic and two fermionic zero

modes, corresponding to $n = 0, \pm 1$ and $m = \pm \frac{1}{2}$. These three zero modes are the infinitesimal action of $\text{OSp}(1|2)$ on the classical solution, and we should quotient the path integral space by this group. This means that in (C.8), we should take the product over $n = 2, 3, 4, \dots$ and $m = \frac{3}{2}, \frac{5}{2}, \frac{7}{2}, \dots$. Using zeta function regularization or an exponential cutoff, one finds the result

$$(C.9) \quad Z_{\text{SJT}}^D(\beta) = e^{S_0} \sqrt{\frac{2}{\pi\beta}} e^{\frac{\pi^2}{\beta}}.$$

The second case we need is the “trumpet” geometry, for which $S = -b^2/2\beta^2$, see [35]. In this case, the only zero mode is the $n = 0$ mode of ε . Inserting this value of S in (C.8) and taking the renormalized product over $n = 1, 2, 3, \dots$ and $m = \frac{1}{2}, \frac{3}{2}, \frac{5}{2}, \dots$, one finds

$$(C.10) \quad Z_{\text{SJT}}^T(\beta, b) = \frac{1}{\sqrt{2\pi\beta}} e^{-\frac{b^2}{4\beta}}.$$

It is also possible to consider the Schwarzian path integral in the trumpet geometry for the case of a Ramond spin structure on the trumpet. The difference is that the fermionic field η is now integrally moded and in particular has a zero-mode, which causes the path integral to vanish. Thus if \tilde{T} is a trumpet with Ramond spin structure, the corresponding Schwarzian path integral is just

$$(C.11) \quad Z_{\text{SJT}}^{\tilde{T}}(\beta, b) = 0.$$

On the other hand, instead of a trumpet with Ramond spin structure, we can consider a disc with a Ramond puncture at the center. The monodromy around a Ramond puncture was described in eqn. (A.22). The important difference between a disc with Ramond puncture and a trumpet with Ramond spin structure is that a certain linear combination of the fermionic generators of $\text{osp}(1|2)$ is a symmetry of a disc with Ramond puncture, since it commutes with the monodromy. In the Schwarzian path integral, unbroken generators of $\text{osp}(1|2)$ are treated as gauge symmetries and the corresponding modes of the Schwarzian multiplet are omitted. In the present case, this means that we omit the zero-mode of the field η . The remaining Schwarzian path integral is extremely simple. The classical action is $S = 0$ (we just set $b = 0$ in the formula $S = -b^2/2\beta^2$ for a trumpet, since geometrically a disc with a puncture is the $b \rightarrow 0$ limit of a trumpet). Moreover, because of unbroken supersymmetry, the 1-loop path integral of the fields ε, η is equal to 1 (these fields are both integrally moded, and their determinants cancel

mode by mode). So the Schwarzian path integral for a disc with Ramond puncture is just

$$(C.12) \quad Z_{\text{SJT}}^{\text{R}}(\beta) = 1.$$

There is no factor of e^{S_0} in eqn. (C.10) or (C.12), since the Euler characteristic of a trumpet or a once-punctured disc is 0.

Appendix D. Mirzakhani's recursion relation and its superanalog

In this appendix, we will review Mirzakhani's derivation of a recursion relation on volumes of moduli spaces of Riemann surfaces, and generalize this derivation to super Riemann surfaces. The underlying super McShane identity has also been proved, for the basic case of a once-punctured torus, by Y. Huang, R. Penner, and A. Zeitlin [59].

D.1. The sum rule

Mirzakhani's sum rule for hyperbolic Riemann surfaces with geodesic boundary generalized an earlier version by McShane for hyperbolic Riemann surfaces with cusps (or punctures) [105]. There is actually a useful short explanation in section 2 of [52], where some of the considerations were generalized to the unorientable case. Here we will consider orientable surfaces only.

We consider a hyperbolic surface Y with a geodesic boundary γ , of length b , on which we will focus. Y may have a set I of additional geodesic boundaries $\gamma_1, \dots, \gamma_n$, of lengths $B = \{b_1, \dots, b_n\}$. We assume that Y is not itself a three-holed sphere. If Y is a three-holed sphere, then the moduli space of hyperbolic structures on Y with given boundary lengths is a point, of volume 1; this will ultimately be the trivial case with which Mirzakhani's recursion relation begins.

Starting at any point $p \in \gamma$, there is a unique inward going geodesic ℓ_p on Y that is orthogonal to γ . We decompose γ as the disjoint union of three sets $\mathcal{A}, \mathcal{B}, \mathcal{C}$, as follows. A point $p \in \gamma$ is in \mathcal{A} if ℓ_p intersects itself, or returns to γ , before meeting any of the other γ_i ; p is in \mathcal{B} if ℓ_p reaches one of the other γ_i before intersecting itself or returning to γ ; and p is in \mathcal{C} if it continues forever in the interior of Y without self-intersection. In case \mathcal{A} or \mathcal{B} , we truncate ℓ_p as soon as it intersects itself or reaches the boundary of Y .

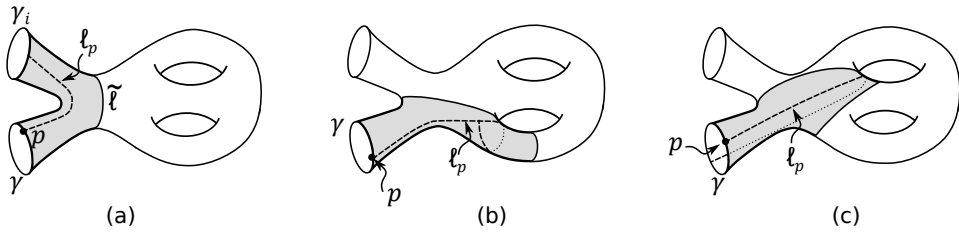


Figure D1: We show three-holed spheres associated to the different fates of a geodesic ℓ_p starting orthogonally at a point p on boundary γ . In (a) $p \in \mathcal{B}_i$, and the geodesic makes it to boundary γ_i . A thickening of ℓ_p, γ, γ_i is topologically the three-holed sphere shown shaded. In (b) and (c), $p \in \mathcal{A}$ and the geodesic either self-intersects or returns to the original boundary. In both cases, a thickening of ℓ_p, γ determines a topological class of three-holed sphere, shown shaded.

A theorem of Birman and Series implies that the set \mathcal{C} is of measure 0; roughly, it is very difficult for a geodesic in a hyperbolic surface to continue indefinitely without self-intersection. Therefore, if we write $\mu(\mathcal{A})$ and $\mu(\mathcal{B})$ for the measures of \mathcal{A} and of \mathcal{B} , then the sum of these measures is the total measure of γ , namely b :

$$(D.1) \quad b = \mu(\mathcal{A}) + \mu(\mathcal{B}).$$

The set \mathcal{B} has an obvious decomposition as the union of subsets \mathcal{B}_i , where $p \in \mathcal{B}_i$ if ℓ_p reaches γ_i first. Thus $\mathcal{B} = \coprod_{i \in I} \mathcal{B}_i$, where \coprod denotes the disjoint union of sets. However, \mathcal{B}_i has a much more subtle and surprising decomposition, as follows. Suppose that $p \in \mathcal{B}_i$, and consider the subset of Y that consists of the union of γ, γ_i , and the segment of ℓ_p that connects them, see fig. D1(a). Thickening this subset slightly, we get a subset \mathcal{S} of Y that is topologically a three-holed sphere. \mathcal{S} has three boundary circles; two of them – namely γ and γ_i – are geodesics. The third is not a geodesic, but by minimizing its length within its homotopy class, we do get a geodesic $\tilde{\ell} \subset Y$. And then γ, γ_i , and $\tilde{\ell}$ are geodesic boundaries of a three-holed sphere $\Lambda \subset Y$. Because we have assumed that Y is not itself a three-holed sphere, $\tilde{\ell}$ is in the interior of Y , and Λ is a proper subset of Y .

Let Υ_i be the set of all three-holed spheres in Y whose boundary consists of γ, γ_i , and an internal geodesic $\tilde{\ell} \subset Y$. In general there are infinitely many choices of $\tilde{\ell}$, so Υ_i – and similarly Υ below – is a (countably) infinite set. We have seen that every $p \in \mathcal{B}_i$ is naturally associated with some $\Lambda \in \Upsilon_i$.

So we have a decomposition

$$(D.2) \quad \mathcal{B} = \coprod_{i \in I, \Lambda \in \Upsilon_i} \mathcal{B}_{i,\Lambda},$$

where $\mathcal{B}_{i,\Lambda}$ consists of points $p \in \mathcal{B}_i$ such that the three-holed sphere that is obtained from the construction that was just described is Λ . Hence

$$(D.3) \quad \mu(\mathcal{B}) = \sum_{i \in I, \Lambda \in \Upsilon_i} \mu(\mathcal{B}_{i,\Lambda}).$$

The set \mathcal{A} has a similar decomposition. If $p \in \mathcal{A}$, we consider the subset of Y that consists of the union of γ with ℓ_p . Thickening this set slightly gives a subset \mathcal{S} of Y that is again topologically a three-holed sphere, see fig. D1(b) and D1(c); remember that ℓ_p is truncated as soon as it intersects itself or returns to γ . \mathcal{S} has a single geodesic boundary, namely γ , and two boundary circles that are not geodesics. By minimizing the lengths of the nongeodesic boundaries in their homotopy classes, we can find geodesics which together with γ bound a three-holed sphere Λ . It might be the case that one of these two additional geodesics is an external boundary γ_i , so that $\Lambda \in \Upsilon_i$. Or it might be the case that both are internal boundaries. Then we say $\Lambda \in \Upsilon$, where Υ is the set of all three-holed spheres in Y whose boundary consists of γ and two internal geodesics. This leads to

$$(D.4) \quad \mu(\mathcal{A}) = \sum_{\Lambda \in \Upsilon} \mu(\mathcal{A}_\Lambda) + \sum_{i \in I, \Lambda \in \Upsilon_i} \mu(\mathcal{A}_{i,\Lambda}),$$

where $\mathcal{A}_\Lambda, \mathcal{A}_{i,\Lambda} \subset \mathcal{A}$ are defined by saying that $p \in \mathcal{A}_\Lambda$ or $p \in \mathcal{A}_{i,\Lambda}$ if ℓ_p is related to $\Lambda \in \Upsilon$ or $\Lambda \in \Upsilon_i$ in the way just described. Combining (D.3) and (D.4), we can write the sum rule (D.1) as

$$(D.5) \quad b = \sum_{\Lambda \in \Upsilon} \mu(\mathcal{A}_\Lambda) + \sum_{i \in I, \Lambda \in \Upsilon_i} \left(\mu(\mathcal{B}_{i,\Lambda}) + \mu(\mathcal{A}_{i,\Lambda}) \right).$$

In this formula, the quantities $\mu(\mathcal{A}_\Lambda)$, $\mu(\mathcal{A}_{i,\Lambda})$, and $\mu(\mathcal{B}_{i,\Lambda})$ have the nice property that they depend only on Λ , and not on anything else about Y . In other words, we can compute them by studying orthogonal geodesics in a three-holed sphere Λ , without worrying about the rest of Y . Suppose that Λ_0 is a three-holed sphere with geodesic boundaries $\gamma, \gamma', \gamma''$ of lengths b, b', b'' . Apart from a set of measure 0, analogous to \mathcal{C} above, we decompose γ as the union of disjoint sets $\mathcal{A}_0, \mathcal{B}'_0$, and \mathcal{B}''_0 , where $p \in \mathcal{A}_0$ if ℓ_p intersects itself or returns to γ before reaching γ' or γ'' , while $p \in \mathcal{B}'_0$ if ℓ_p reaches γ' first, and

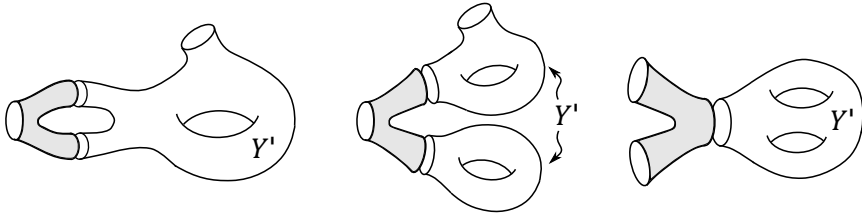


Figure D2: We show three ways of building a $g = 2$ surface Y with two boundaries by gluing a three-holed sphere onto another hyperbolic surface Y' , possibly disconnected. These three types of gluing correspond respectively to the three types of terms on the RHS of (D.11).

$p \in \mathcal{B}_0''$ if ℓ_p reaches γ'' first. We set $\mathsf{T}(b, b', b'') = \mu(\mathcal{B}_0'')$, so that by symmetry $\mu(\mathcal{B}_0'') = \mathsf{T}(b, b'', b')$. Similarly we let $\mathsf{D}(b, b', b'') = \mu(\mathcal{A}_0)$. We will compute $\mathsf{T}(b, b', b'')$ in section D.4. This will determine $\mathsf{D}(b, b', b'')$ because of the sum rule

$$(D.6) \quad b = \mathsf{D}(b, b', b'') + \mathsf{T}(b, b', b'') + \mathsf{T}(b, b'', b').$$

Every three-holed sphere Λ is isomorphic to Λ_0 for some values of b, b', b'' , so for any Λ , the quantities $\mu(\mathcal{A}_\Lambda)$, $\mu(\mathcal{A}_{i,\Lambda})$, and $\mu(\mathcal{B}_{i,\Lambda})$ in eqn. (D.5), which measure the sets of orthogonal geodesics with a specified behavior inside Λ , can be expressed in terms of the functions $\mathsf{T}(b, b', b'')$ and $\mathsf{D}(b, b', b'')$. In fact, $\mu(\mathcal{A}_\Lambda) = \mathsf{D}(b, b', b'')$, while $\mu(\mathcal{B}_{i,\Lambda}) = \mathsf{T}(b, b_i, b')$, $\mu(\mathcal{A}_{i,\Lambda}) = \mathsf{D}(b, b_i, b')$. So a somewhat more explicit version of eqn. (D.5) is

$$(D.7) \quad b = \sum_{\Lambda \in \Upsilon} \mathsf{D}(b, b', b'') + \sum_{i \in I, \Lambda \in \Upsilon_i} \left(\mathsf{T}(b, b_i, b') + \mathsf{D}(b, b_i, b') \right)$$

$$(D.8) \quad = \sum_{\Lambda \in \Upsilon} \mathsf{D}(b, b', b'') + \sum_{i \in I, \Lambda \in \Upsilon_i} \left(b - \mathsf{T}(b, b', b_i) \right).$$

D.2. The recursion relation

As preparation for explaining Maryam Mirzakhani’s recursion relation for volumes, we will describe how one might proceed if there were no need to divide by the mapping class group.

Let Y be a hyperbolic surface with boundary lengths b and b_i , $i \in I$. Suppose that Y can be built by gluing a three-holed sphere Λ of boundary lengths b, b', b'' onto another hyperbolic surface Y' with boundary lengths b', b'' , and b_i , $i \in I$, see fig. D2. If Y has genus g with $n + 1$ boundaries,

then Y' might be, for example, a surface of genus $g - 1$ with $n + 2$ boundaries (Y' can also be disconnected, as shown in the figure). Looking back to eqn. (3.44), we see that the torsions (or volume forms) of Y and Y' are related by $\tau_Y = db'd\rho'db''d\rho''\tau_{Y'}$. If the mapping class group were not relevant, this relation between the volume forms would lead after integration to a relation between the volumes:

$$(D.9) \quad V_g(b, B) \stackrel{?}{=} \frac{1}{2} \int db'd\rho'db''d\rho'' V_{g-1}(b', b'', B).$$

Here B represents the collection b_1, \dots, b_n , and the factor of $1/2$ takes into account the symmetry under exchange of b' and b'' , to avoid double-counting. We can make this formula more explicit by noting that b', b'' run over the half-lines $[0, \infty)$, while ρ', ρ'' run over $0 \leq \rho' \leq b', 0 \leq \rho'' \leq b''$. Moreover, the volumes do not depend on ρ', ρ'' , so the integration over those variables just gives a factor $b'b''$. Thus the formula would reduce to

$$(D.10) \quad V_g(b, B) \stackrel{?}{=} \frac{1}{2} \int_0^\infty b'db' b''db'' V_{g-1}(b', b'', B).$$

The analog of this formula for a compact group such as $SU(2)$ is actually correct, as discussed for example in section 2 of [72]. However, for $SL(2, \mathbb{R})$, one needs to divide by the mapping class group. Eqn. (D.9) is wrong because the mapping class group of Y is larger than the mapping class group of Y' , or to put it differently, because modular invariance would force us in eqn. (D.10) to sum over contributions of infinitely many choices of the three-holed sphere Λ . This sum will give a divergent factor, so clearly the formula (D.10) is not correct. (In fact, the volumes $V_g(b_1, \dots, b_s)$ are polynomials in b_1^2, \dots, b_s^2 , so the integral on the right hand side of eqn. (D.10) will diverge.)

If one could find a function $f(b, b', b'')$ with the property that its sum over all possible choices of Λ is 1, then one could find a correct version of eqn. (D.10) by inserting in the definition $V_g(b, B) = \int_{\mathcal{M}_{g,n+1}} \tau_Y$ a factor of $1 = \sum_\Lambda f(b, b', b'')$. Then, instead of summing over Λ and dividing by the mapping class group of Y , it would be equivalent to pick a particular Λ and divide by the mapping class group of Y' . The upshot would be a recursion relation rather like eqn. (D.10), but with an extra factor $f(b, b', b'')$ in the integral on the right hand side.

There is no function $f(b, b', b'')$ that has quite the property we stated, but as observed by Mirzakhani, the sum rule of eqn. (D.7) can be used similarly. On the right hand side of eqn. (D.7), the summation over Λ includes a sum over all topological ways of obtaining Y by gluing a three-holed sphere Λ

onto some other surface Y' (with the constraint that one of the boundaries of Λ is a specified boundary of Y). So a recursion relation derived from this sum rule involves a sum over all topological types of gluing. It reads⁷²

$$\begin{aligned}
 \text{(D.11)} \quad bV_{g,n+1}(b, B) &= \frac{1}{2} \int_0^\infty b' db' b'' db'' D(b, b', b'') \\
 &\quad \times \left(V_{g-1}(b', b'', B) + \sum_{\text{stable}} V_{h_1}(b', B_1) V_{h_2}(b'', B_2) \right) \\
 &\quad + \sum_{k=1}^{|B|} \int_0^\infty b' db' \left(b - \Upsilon(b, b', b_k) \right) V_g(b', B \setminus b_k).
 \end{aligned}$$

This is Mirzakhani’s recursion relation [37], which was interpreted by Eynard and Orantin [36] in terms of topological recursion for matrix ensembles. For a certain supersymmetric matrix ensemble, we deduced in eqn. (5.42) a recursion relation with the same structure. We will aim to find a superanalog of Mirzakhani’s derivation to account for that result.

D.3. More on hyperbolic geometry and three-holed spheres

As preparation for computing $\Upsilon(b, b', b'')$, and for understanding the superanalog of Mirzakhani’s derivation, we need some more details on hyperbolic geometry.

Let $z = x + iy$ with real x, y . The group $\text{SL}(2, \mathbb{R})$ acts on the upper half-plane $y > 0$, leaving fixed the hyperbolic metric

$$\text{(D.12)} \quad ds^2 = \frac{dzd\bar{z}}{\text{Im}^2 z} = \frac{dx^2 + dy^2}{y^2}.$$

We call this half-plane H . The line $x = 0$ plus a point at infinity constitute the “conformal boundary” of H , which we will call ∂H . As is usual in holographic duality, the conformal boundary is “at infinity,” not contained in H . Including the point at infinity, ∂H is a circle topologically.

H is complete, so a geodesic in H propagates for infinite distance in each direction, but a geodesic in H has at each end a “virtual endpoint” in

⁷² There is a subtlety in terms involving $V_1(b)$. Because of the \mathbb{Z}_2 symmetry of the torus with one hole, when we glue in such a surface, we should restrict the twist to be between zero and $b/2$ instead of between zero and b . In order to avoid awkward factors of two that result from this, we are simply going to define $V_1(b)$ to be one half of the true moduli space volume. Then we can integrate the twist from zero to b as usual.

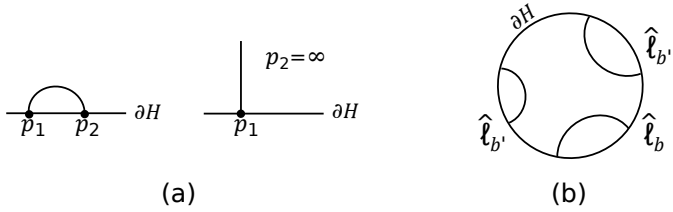


Figure D3: In (a) we view ∂H as the real line, and we show a geodesic with virtual endpoints at finite locations in ∂H , as well as one with a virtual endpoint at infinity. In (b) we view ∂H as S^1 , and we show a non-nested configuration of three geodesics.

∂H . Conversely, any two points in ∂H are connected by a unique geodesic in H . These geodesics are easily described in terms of the Euclidean metric $ds_E^2 = dx^2 + dy^2$ on H . If p_1, p_2 are finite points in ∂H , then the geodesic between them is a semicircle, centered at the midpoint of the interval $[p_1, p_2]$, and with diameter such that it passes through p_1 and p_2 . If, say, p_2 is the point at infinity, then the geodesic between p_1 and p_2 is a vertical straight line from p_1 . See fig. D3(a).

The conformal boundary ∂H can be understood as a copy of \mathbb{RP}^1 , with homogeneous real coordinates u, v . $SL(2, \mathbb{R})$ acts on \mathbb{RP}^1 in the natural way

$$(D.13) \quad \begin{pmatrix} u \\ v \end{pmatrix} \rightarrow \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix}, \quad ad - bc = 1.$$

The corresponding action on $x = u/v$ is familiar:

$$(D.14) \quad x \rightarrow \frac{ax + b}{cx + d}.$$

The point at $x = \infty$ is, of course, the point in \mathbb{RP}^1 with $(u, v) = (1, 0)$.

The action of $SL(2, \mathbb{R})$ on H is described by the same formula, but with x replaced by a complex variable z (constrained to have $\text{Im } z > 0$):

$$(D.15) \quad z \rightarrow \frac{az + b}{cz + d}.$$

Now we will discuss some details about three-holed spheres that were not needed in the body of the paper. If Y is a three-holed hyperbolic sphere, with geodesic boundaries, then its universal cover \hat{Y} is a region in H bounded by three geodesics.

Let us work this out for the case studied in section 3.4 of a three-holed sphere with monodromies

$$(D.16) \quad U_0 = \delta_b \begin{pmatrix} e^{b/2} & \kappa \\ 0 & e^{-b/2} \end{pmatrix}, \quad V_0 = \delta_{b'} \begin{pmatrix} e^{-b'/2} & 0 \\ 1 & e^{b'/2} \end{pmatrix},$$

and $W_0 = V_0^{-1}U_0^{-1}$. (We write b, b', b'' for what in section 3.4 were a, b, c .) The length parameters of U_0, V_0 , and W_0 are b, b' , and a parameter b'' that satisfies

$$(D.17) \quad 2 \cosh \frac{b''}{2} = -2 \cosh \frac{b - b'}{2} - \kappa$$

or

$$(D.18) \quad \kappa = -2 \cosh \frac{b - b'}{2} - 2 \cosh \frac{b''}{2}.$$

Eqn. (D.17) does not determine the sign of b'' , and we can pick $b'' > 0$ as a convention.

Let ℓ_b be the the boundary of Y whose monodromy is U_0 . Then ℓ_b lifts on H to a U_0 -invariant geodesic $\widehat{\ell}_b$. For $\widehat{\ell}_b$ to be U_0 -invariant is equivalent to saying that its virtual endpoints on ∂H are U_0 -invariant. A hyperbolic element of $SL(2, \mathbb{R})$, such as U_0 , has precisely two fixed points on $\partial H \cong \mathbb{RP}^1$. These fixed points correspond to the eigenvectors of the 2×2 matrix U_0 , since if

$$(D.19) \quad U_0 \begin{pmatrix} u \\ v \end{pmatrix} = \lambda \begin{pmatrix} u \\ v \end{pmatrix},$$

for some λ , this means that U_0 leaves invariant the point in \mathbb{RP}^1 that corresponds to $\begin{pmatrix} u \\ v \end{pmatrix}$. The endpoints of $\widehat{\ell}_b$ are thus the points on \mathbb{RP}^1 that correspond to eigenvectors of U_0 .

The eigenvectors of U_0 are $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} -\frac{\kappa}{2 \sinh \frac{b}{2}} \\ 1 \end{pmatrix}$. So the endpoints of the geodesic $\widehat{\ell}_b$ corresponding to U_0 are the points with $x = \infty$ and $x = -\kappa/(2 \sinh \frac{b}{2})$, respectively.

Similarly, the eigenvectors of V_0 are $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ and $\begin{pmatrix} -2 \sinh \frac{b'}{2} \\ 1 \end{pmatrix}$. So the endpoints of the geodesic $\widehat{\ell}_{b'}$ corresponding to V_0 are the point at $x = 0$ and the point at $x = -2 \sinh \frac{b'}{2}$.

The endpoints of the third geodesic $\widehat{\ell}_{b''}$, similarly correspond to the eigenvectors of W_0 , which can be found using the explicit formula (3.22). The endpoints of $\widehat{\ell}_{b''}$ are at $x = -(e^{b'/2} + e^{(b\pm b'')/2})$.

At this point, we can verify that the matrices U_0, V_0, W_0 really are the monodromies of a three-holed sphere Y with hyperbolic metric (and not of a flat $SL(2, \mathbb{R})$ connection in a different component of the moduli space). If such a Y does exist, then, after we lift to the upper half-plane, the geodesics $\widehat{\ell}_b, \widehat{\ell}_{b'}$, and $\widehat{\ell}_{b''}$ are nonintersecting and moreover they are not “nested”; the cyclic ordering of the endpoints of the three geodesics on $\partial H = S^1$ must be as shown in fig. D3(b), up to a possible permutation of the three geodesics. A sufficient condition is to take b, b' both positive. (As noted above, there is no loss of generality to assume that also $b'' > 0$.) From the above formulas, the endpoints of $\widehat{\ell}_b$ are more positive than those of $\widehat{\ell}_{b'}$ which in turn are more positive than those of $\widehat{\ell}_{b''}$. So the endpoints are arranged appropriately and we do get a hyperbolic metric on a three-holed sphere for any positive boundary lengths b, b', b'' .

D.4. Computation

We will now compute the function $T(b, b', b'')$ that was introduced in appendix D.1. Consider a three-holed sphere Λ with geodesic boundaries $\gamma_b, \gamma_{b'}, \gamma_{b''}$ of lengths b, b', b'' . The universal cover of Λ can be identified as a region in the upper half-plane H bounded by geodesics $\widehat{\gamma}_b, \widehat{\gamma}_{b'}$, and $\widehat{\gamma}_{b''}$, which are obtained respectively by unwrapping $\gamma_b, \gamma_{b'}, \gamma_{b''}$.

Up to an $SL(2, \mathbb{R})$ transformation, we can identify $\widehat{\gamma}_b$ as a vertical straight line in H located at, say, $x = x_0$. In the same description, $\widehat{\gamma}_{b'}$ will be a (Euclidean) semi-circle in H with endpoints at, say, x_1, x_2 . Since $\widehat{\gamma}_b$ and $\widehat{\gamma}_{b'}$ do not intersect, we can assume (up to relabeling and reversal of orientation) that the endpoints are arranged with $x_2 < x_1 < x_0$ (fig. D4). We want to compute the measure of the set $\mathcal{B}' \subset \widehat{\gamma}_b$ defined by saying that $x \in \mathcal{B}'$ if the orthogonal geodesic ℓ_p to $\widehat{\gamma}_b$ at p meets $\widehat{\gamma}_{b'}$. From the figure, it is fairly easy to see that \mathcal{B}' is an open interval. The endpoints of the closure of this interval are points $s_1, s_2 \in \widehat{\gamma}_b$ with the property that ℓ_{s_1} and ℓ_{s_2} do not actually intersect $\widehat{\gamma}_{b'}$, but rather have an endpoint on ∂H that coincides with one of the endpoints of $\widehat{\gamma}_{b'}$.

From this description, and the fact that geodesics in H are Euclidean semicircles, it follows that the (x, y) coordinates of the points s_1, s_2 are $(x_0, x_0 - x_1)$ and $(x_0, x_0 - x_2)$. $\mu(\mathcal{B}')$ is simply the distance between these points in the hyperbolic metric of eqn. (D.12). A simple integration gives

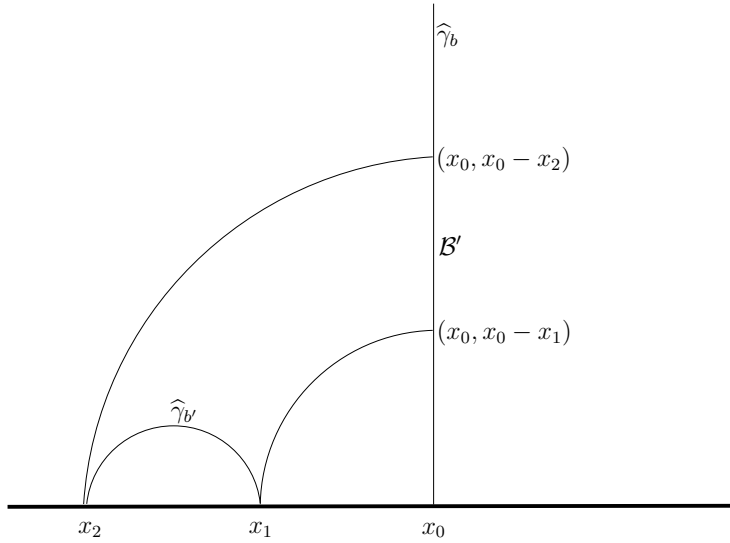


Figure D4: Here the geodesic $\hat{\gamma}_b$ is the vertical straight line $x = x_0$ in H and $\hat{\gamma}_{b'}$ is a semicircle with endpoints at $(x, y) = (x_1, 0)$ and $(x_2, 0)$. \mathcal{B}' consists of points $p \in \hat{\gamma}_b$ such that the orthogonal geodesic ℓ_p to $\hat{\gamma}_b$ at p intersects $\hat{\gamma}_{b'}$. \mathcal{B}' is an open interval; the endpoints of its closure are the points p such that ℓ_p has an endpoint in common with $\hat{\gamma}_{b'}$. These are the points $(x, y) = (x_0, x_0 - x_1)$ and $(x, y) = (x_0, x_0 - x_2)$. The distance between those two points in the metric on H is $\log(x_0 - x_2)/(x_0 - x_1)$.

$\mu(\mathcal{B}') = \log(x_0 - x_2)/(x_0 - x_1)$. (To express this more invariantly, the ratio $(x_0 - x_2)/(x_0 - x_1)$ is a cross-ratio of the four points x_0, x_1, x_2, ∞ , where ∞ is the second endpoint of $\hat{\gamma}_b$.)

To turn this into a formula for $\mathbb{T}(b, b', b'')$, we use the values of x_0, x_1, x_2 that were computed in appendix D.3 for a three-holed sphere with boundary lengths b, b', b'' . Thus $x_0 = -\kappa/2 \sinh \frac{b}{2} = (\cosh \frac{b''}{2} + \cosh \frac{b-b'}{2})/\sinh \frac{b}{2}$, $x_1 = 0$, $x_2 = -2 \sinh \frac{b'}{2}$. Hence

$$(D.20) \quad \mathbb{T}(b, b', b'') = \log \frac{x_0 - x_2}{x_0 - x_1} = \log \frac{\cosh \frac{b''}{2} + \cosh \frac{b+b'}{2}}{\cosh \frac{b''}{2} + \cosh \frac{b-b'}{2}},$$

which is Mirzakhani’s result.

In using this formula, Mirzakhani differentiated the recursion relation (D.11) with respect to b , so as to replace the functions D and T with trigonometric functions. We will see that in the supersymmetric case, the integration over the odd moduli automatically gives us trigonometric functions.

D.5. Some supergeometry

To generalize this computation to the case with $\mathcal{N} = 1$ supersymmetry, we will have to be familiar with some aspects of the supersymmetric analog \widehat{H} of the upper half-plane. Some old and new references on this material are [97, 106–109].

In section 3.5.1, we defined $\text{OSp}(1|2)$ as the symmetry group of a copy of $\mathbb{R}^{2|1}$, parametrized by bosonic and fermionic coordinates $u, v|\theta$, that preserve the symplectic form $\widehat{\omega} = dudv + \frac{1}{2}d\theta^2$. If we view $u, v|\theta$ as homogeneous coordinates, then they parametrize a copy of $\mathbb{RP}^{1|1}$. This will be the conformal boundary $\partial\widehat{H}$ of the supersymmetric analog \widehat{H} of the upper half-plane.

$\mathbb{RP}^{1|1}$ can be parametrized by ordinary coordinates $x|\theta$, defined by $x = u/v$, $\theta = \theta/v$, along with a divisor at infinity (which corresponds to the case $v = 0$). The Lie algebra $\text{osp}(1|2)$ is represented in these coordinates as follows:⁷³

$$\begin{aligned}
 e &= \partial_x \\
 h &= 2x\partial_x + \theta\partial_\theta \\
 f &= -x^2\partial_x - x\theta\partial_\theta \\
 q_1 &= -\partial_\theta + \theta\partial_x \\
 q_2 &= x(\partial_\theta - \theta\partial_x).
 \end{aligned}
 \tag{D.21}$$

As coordinates on $\partial\widehat{H}$, x and θ are real. To describe \widehat{H} and the action on it of $\text{osp}(1|2)$, we simply complexify these formulas. Thus we replace the real even and odd coordinates $x|\theta$ with complex even and odd coordinates $z|\vartheta$, with z constrained by $\text{Im } z > 0$. The action of $\text{osp}(1|2)$ is given by the

⁷³These vector fields actually generate the “opposite” Lie superalgebra to the algebra of matrices in eqns. (3.62) and (3.63). The opposite superalgebra is defined by reversing the signs of commutators and leaving the anticommutators unchanged. (The superalgebra $\text{osp}(1|2)$ and its opposite are equivalent under $(e, f, h, q_1, q_2) \rightarrow (-f, -e, h, q_2, q_1)$.)

same formulas as before, with $x|\theta$ replaced by $z|\vartheta$:

$$\begin{aligned}
 e &= \partial_z \\
 h &= 2z\partial_z + \vartheta\partial_\vartheta \\
 f &= -z^2\partial_z - z\vartheta\partial_\vartheta \\
 q_1 &= -\partial_\vartheta + \vartheta\partial_z \\
 q_2 &= z(\partial_\vartheta - \vartheta\partial_z).
 \end{aligned}
 \tag{D.22}$$

For our purposes here,⁷⁴ a super Riemann surface is a real supermanifold of dimension 2|2. The complex conjugate coordinates $\bar{z}|\bar{\vartheta}$ transform under $\text{osp}(1|2)$ in the same way as $z|\theta$; one can simply take the complex conjugate of all formulas in eqn. (D.22).

One point about this definition is that the bulk supermanifold \widehat{H} is of real dimension 2|2, while its conformal “boundary” is of dimension 1|1, and thus is of codimension 1|1, as opposed to the codimension 1|0 of an ordinary boundary. That might come as a surprise, but this behavior is actually typical when holographic dualities are described in superspace. Dimension 2|2 is appropriate to realize $\text{OSp}(1|2)$ as a supergroup of symmetries in bulk, and dimension 1|1 is appropriate to realize $\text{OSp}(1|2)$ as a supergroup of superconformal symmetries on the boundary. The relation between \widehat{H} and $\partial\widehat{H}$ was described from a different perspective in [109].

Now, recall that in the bosonic case, any two points p_1 and p_2 in ∂H , say the points with $x = x_1$ or with $x = x_2$, are the endpoints of a unique geodesic γ in H . Using the embedding of H in \widehat{H} , we embed γ in \widehat{H} and declare it to be the geodesic connecting the points $x_1|0$ and $x_2|0$. To generalize this to boundary points with nonzero θ , we just note that any desired pair $x_1|\theta_1$ and $x_2|\theta_2$ can be reached from $x_1|0$ and $x_2|0$ by an $\text{OSp}(1|2)$ transformation g (which moreover is unique up to a symmetry of γ). We declare $\tilde{\gamma} = g(\gamma)$ to be the geodesic in \widehat{H} with endpoints $x_1|\theta_1$ and $x_2|\theta_2$. It is possible to work out

⁷⁴In superstring theory, a super Riemann surface must be understood in a more subtle way, not simply as a real supermanifold. This is because the holomorphic and antiholomorphic spin structures and odd moduli vary independently (Type II superstrings) or because the supersymmetric structure is purely holomorphic (the heterotic string). See for example section 5 of [103]. Our present problem is more straightforward, because as long as we are on an orientable two-manifold, all supermanifolds we encounter can be understood as real supermanifolds. On an unorientable two-manifold, one has to extend $\text{OSp}(1|2)$ in such a way that the odd variables no longer carry a real structure (see section 3.5.1), and then some of the issues that arise in superstring theory do become relevant.

an explicit formula for $\tilde{\gamma}$, but we will not need it; our explicit calculations will only involve the endpoints of geodesics, and special geodesics at $\vartheta = \bar{\vartheta} = 0$.

In the bosonic world, if γ_1 and γ_2 are geodesics with endpoints at $x = a_1$ and $x = a_2$, respectively, then we say that γ_1 and γ_2 have a common endpoint if $a_1 = a_2$. In the supersymmetric case, if γ_1 and γ_2 have respective endpoints $a_1|\theta_1$ and $a_2|\theta_2$, the naive generalization of this is to require $a_1 = a_2$, $\theta_1 = \theta_2$. This is too strong, however. It is usually more useful to impose only the single condition

$$(D.23) \quad a_1 - a_2 - \theta_1\theta_2 = 0,$$

which for most purposes is the closest analog of “having a common endpoint” in the bosonic case.

An alternative way to state this condition is as follows. The vector field $D_\theta = \partial_\theta + \theta\partial_x$ is not $\text{OSp}(1|2)$ -invariant, but it is $\text{OSp}(1|2)$ invariant up to a transformation $D_\theta \rightarrow e^{f(x|\theta)}D_\theta$. (More specifically, D_θ transforms as a superconformal primary of weight $-1/2$, a fact that is used in constructing superconformally invariant Lagrangians.) The orbits generated by D_θ are invariant under $D_\theta \rightarrow e^f D_\theta$, so they are superconformally invariant. Concretely the orbit containing the point $x_1|\theta_1$ is parametrized by an odd variable α as follows:

$$(D.24) \quad x|\theta = x_1 + \alpha\theta_1|\theta_1 + \alpha.$$

These orbits are said to be the “leaves” of an “unintegrable distribution.” For our purposes, the relevant point is that the condition $a_1 - a_2 - \theta_1\theta_2 = 0$ is equivalent to saying that $a_1|\theta_1$ and $a_2|\theta_2$ are contained in the same leaf. Indeed, if $a_1 - a_2 - \theta_1\theta_2 = 0$, then the leaf through the point $a_1|\theta_1$ contains the point $a_2|\theta_2$, as we see by setting $\alpha = \theta_2 - \theta_1$ in eqn. (D.24).

A similar idea is actually built into the naive statement “ Y is a super Riemann surface with geodesic boundary.” A super Riemann surface in the sense of our present discussion has real dimension $2|2$. Its boundary should have codimension $1|0$, so should be a supermanifold of dimension $1|2$. But a geodesic has dimension $1|0$. The point is that the vector fields on \hat{H} defined by $D_\vartheta = \partial_\vartheta + \vartheta\partial_z$ and $D_{\bar{\vartheta}} = \partial_{\bar{\vartheta}} + \bar{\vartheta}\partial_{\bar{z}}$ generate a superconformally invariant distribution whose leaves are of dimension $0|2$. Hence a geodesic γ (or similarly any generic curve of dimension $1|0$) has a canonical dimension $1|2$ thickening, consisting of all the leaves that pass through γ . It is this thickening that is really the boundary of a super Riemann surface.

In H , which is of dimension 2 , it is natural for two geodesics γ_1 and γ_2 , each of which has codimension 1 , to intersect. (Of course, not all pairs of

geodesics in H do intersect.) What is a good analog in \widehat{H} of the statement that two geodesics in H intersect? Since geodesics in \widehat{H} have codimension $1|2$ while \widehat{H} has dimension $2|2$, it is nongeneric for two geodesics in \widehat{H} to have a point in common. The useful notion of “intersection of geodesics” in a super Riemann surface is to say that two geodesics γ_1 and γ_2 intersect if the thickening of γ_1 , in the sense described in the last paragraph, intersects γ_2 . This condition can be stated more symmetrically by saying that the distribution generated by $D_{\mathfrak{g}}$ and $D_{\overline{\mathfrak{g}}}$ has a leaf that intersects both γ_1 and γ_2 in the naive sense.

Finally, we need to know what it means for two geodesics in \widehat{H} to be orthogonal. If a geodesic $\gamma \subset \widehat{H}$ actually lies in $H \subset \widehat{H}$, and p is a point in γ , then by the geodesic ℓ_p that is orthogonal to γ at p , we just mean the geodesic in $H \subset \widehat{H}$ that is orthogonal to γ at p , in the classical sense. Any case can be reduced to this case by an $\mathrm{OSp}(1|2)$ transformation, since by definition any geodesic in \widehat{H} can be mapped to H by an $\mathrm{OSp}(1|2)$ transformation.

D.6. The supersymmetric recursion relation

Now we consider the same setting as in appendix D.1, but in the supersymmetric case. Thus, Y is a super Riemann surface with a specified geodesic boundary γ of length b , and additional geodesic boundaries γ_i , $i \in I$, of lengths b_i . γ has a natural Riemannian measure, since its lift $\widehat{\gamma}$ to \widehat{H} is an orbit of an $\mathrm{SO}(1, 1)$ subgroup of $\mathrm{OSp}(1|2)$. The total measure of $\widehat{\gamma}$ is b .

For $p \in \gamma$, let ℓ_p be the normal to γ at p . Just as in the bosonic case, we decompose γ as a union of disjoint subsets \mathcal{A} , \mathcal{B} , and \mathcal{C} according to whether ℓ_p first intersects itself or returns to γ ; first intersects one of the γ_i ; or does neither. The last case corresponds again to a set of measure 0, and in the first two cases, just as before,⁷⁵ ℓ_p is contained in a distinguished three-holed sphere Λ one of whose geodesic boundaries is γ .

⁷⁵Upon reducing modulo the odd variables (the fermionic coordinates of Y and also the fermionic moduli), Y reduces to a bosonic Riemann surface Y_0 and the same argument as before associates to ℓ_p a three-holed sphere $\Lambda_0 \subset Y_0$ with geodesic boundaries γ_a (one of which is the original γ). Picking any map $i : Y_0 \rightarrow Y$ that reduces to the identity modulo the odd variables, we embed the γ_a in Y as closed loops $i(\gamma_a)$, which reduce modulo the odd variables to the original geodesics γ_a . The $i(\gamma_a)$ are in general not geodesics, but in the homotopy class of each of them (and differing from it only by nilpotent terms, proportional to the odd variables) is a unique geodesic. Indeed, the monodromy of the hyperbolic flat $\mathrm{OSp}(1|2)$ connection around any of the γ_a is a hyperbolic element $h_a \in \mathrm{OSp}(1|2)$. Pass to \widehat{H} by replacing Y with its universal cover. A hyperbolic element of $\mathrm{OSp}(1|2)$, such as h_a , leaves fixed a unique geodesic in \widehat{H} , namely the geodesic that connects its two fixed points in $\mathbb{RP}^{1|1}$. This descends to the desired geodesic in Y . The three geodesics obtained

Thus after decomposing \mathcal{A} and \mathcal{B} as in (D.3) and (D.4), we have the analog of eqn. (D.5):

$$(D.25) \quad b = \sum_{\Lambda \in \Upsilon} \mu(\mathcal{A}_\Lambda) + \sum_{i \in I, \Lambda \in \Upsilon_i} \left(\mu(\mathcal{B}_{i,\Lambda}) + \mu(\mathcal{A}_\Lambda) \right).$$

As before, the measures $\mu(\mathcal{A}_\Lambda)$ and $\mu(\mathcal{B}_{i,\Lambda})$ that appear in these formulas depend only on Λ and not on anything else about Y . Thus they come from universal functions. The main difference is that now a three-holed sphere has two odd moduli as well as the three boundary lengths, and we will have to take these into account.

Let Λ_0 be a three-holed sphere with geodesic boundaries $\gamma, \gamma', \gamma''$ of lengths b, b', b'' , and with odd moduli α, β . For fixed b, b', b'' , the odd variables α and β parametrize a supermanifold $\mathcal{M}_{b,b',b''}$ of dimension $0|2$. As before, γ is the union of sets $\mathcal{A}_0, \mathcal{B}'_0$, and \mathcal{B}''_0 (plus a set of measure 0) defined by the behavior of orthogonal geodesics. Writing $\widehat{D}(b, b', b'' | \alpha, \beta)$, $\widehat{T}(b, b', b'' | \alpha, \beta)$, and⁷⁶ $\widehat{U}(b, b', b'' | \alpha, \beta)$ for the measures of $\mathcal{A}_0, \mathcal{B}'_0$, and \mathcal{B}''_0 , we get a sum rule just like eqn. (D.6), with new universal functions:

$$(D.26) \quad b = \widehat{D}(b, b', b'' | \alpha, \beta) + \widehat{T}(b, b', b'' | \alpha, \beta) + \widehat{U}(b, b', b'' | \alpha, \beta).$$

Just as in eqn. (D.11), inserting this identity in the definition of the volume as an integral over the supermoduli space leads to a recursion relation that involves a sum over all ways that Y can be built by gluing a three-holed sphere Λ to some other super Riemann surface Y' . The main difference is that after specifying the boundary lengths b, b', b'' of Λ , there is still a moduli space $\mathcal{M}_{b,b',b''}$ of flat connections on Λ over which we have to integrate. We will describe in section D.7 the integration measure μ for this integral. Including some more trivial factors that are explained in a moment,

this way (one of which is γ) are the boundaries of the desired three-holed sphere $\Lambda \subset Y$.

⁷⁶ \widehat{U} is obtained from \widehat{T} by exchanging b' with b'' and suitably transforming α and β . We will soon integrate out α and β , so we avoid writing explicit formulas as this stage.

the analog of eqn. (D.11) is then

$$\begin{aligned}
 \text{(D.27)} \quad bV_g(b, B) &= \frac{1}{2} \int_0^\infty b' db' b'' db'' \int_{\mathcal{M}_{b, b', b''}} d\mu \widehat{D}(b, b', b'' | \alpha, \beta) \\
 &\quad \times \left(V_{g-1}(b', b'', B) + \sum_{\text{stable}} V_{h_1}(b', B_1) V_{h_2}(b'', B_2) \right) \\
 &\quad + 2 \sum_{k=1}^{|B|} \int_0^\infty b' db' \\
 &\quad \times \int_{\mathcal{M}_{b, b_k, b'}} d\mu \left(b - \widehat{U}(b, b_k, b' | \alpha, \beta) \right) V_g(b', B \setminus b_k).
 \end{aligned}$$

In the sum over spin structures, we may or may not include a factor $(-1)^\zeta$. However, as explained in section 5.2, the more interesting case is that we do include such a factor (otherwise the volumes are expected to all vanish). We have written the formula for this case. With a factor $(-1)^\zeta$ included, if any of the circles on which we glue Λ to Y' has a spin structure of Ramond type, then the sum over the spin structure “orthogonal” to that circle vanishes, as we learned in section 2.4.3. So we assume that all of those circles have NS spin structure. Similarly, we assume the same for all external boundaries. A factor of 2 in the term proportional to $b - \widehat{U}$ has the following origin. We define the volumes to include a sum over spin structures, which includes a sum over spin structures “orthogonal” to the boundary components (that is, the spin bundle is trivialized on ∂Y , and spin structures are considered equivalent only if there is an equivalence between them that respects this trivialization). In the gluing that leads to the $b - \widehat{U}$ term in eqn. (D.27), but not in the gluing that leads to the \widehat{D} term, the spin structure sum on the left has an extra factor of 2 compared to the spin structure sum on the right. We compensate for this with an explicit factor of 2 multiplying the $b - \widehat{U}$ term.

To get a formula for the volumes in purely bosonic terms, we should integrate over $\mathcal{M}_{b, b', b''}$. We define

$$\begin{aligned}
 \int_{\mathcal{M}_{b, b', b''}} d\mu \widehat{D}(b, b', b'' | \alpha, \beta) &= \mathcal{D}(b, b', b'') \\
 \text{(D.28)} \quad \int_{\mathcal{M}_{b, b', b''}} d\mu \widehat{T}(b, b', b'' | \alpha, \beta) &= \mathcal{T}(b, b', b'').
 \end{aligned}$$

Once we integrate over the odd variables, $\widehat{\mathcal{T}}$ and $\widehat{\mathcal{U}}$ differ only by exchange of b' and b'' , so

$$(D.29) \quad \int_{\mathcal{M}_{b,b',b''}} d\mu \widehat{\mathcal{U}}(b, b', b'' | \alpha, \beta) = \mathcal{T}(b, b'', b')$$

The supersymmetric recursion relation then reduces to

$$(D.30) \quad bV_g(b, B) = \frac{1}{2} \int_0^\infty b' db' b'' db'' \mathcal{D}(b, b', b'') \\ \times \left(V_{g-1}(b', b'', B) + \sum_{\text{stable}} V_{h_1}(b', B_1) V_{h_2}(b'', B_2) \right) \\ - 2 \sum_{k=1}^{|B|} \int_0^\infty b' db' \mathcal{T}(b, b', b_k) V_g(b', B \setminus b_k).$$

In writing this equation, we used that the moduli space $\mathcal{M}_{b,b',b''}$ has zero volume, like all of the genus 0 moduli spaces, so the integral over this moduli space of a constant function vanishes:

$$(D.31) \quad \int_{\mathcal{M}_{b,b',b''}} d\mu \cdot b = 0.$$

This fact has another useful implication. When we integrate the sum rule (D.25) over $\mathcal{M}_{b,b',b''}$, we learn that

$$(D.32) \quad 0 = \mathcal{D}(b, b', b'') + \mathcal{T}(b, b', b'') + \mathcal{T}(b, b'', b').$$

Therefore, to make eqn. (D.30) explicit, it suffices to compute $\mathcal{T}(b, b', b'')$.

D.7. Final calculations

To compute the function $\widehat{\mathcal{T}}$, we consider a three-holed sphere Λ with the monodromies of eqn. (3.76), namely

$$(D.33) \quad U_0 = \delta_b \left(\begin{array}{cc|c} e^{b/2} & \kappa & 0 \\ 0 & e^{-b/2} & 0 \\ \hline 0 & 0 & \delta_b \end{array} \right) \exp(\xi q_1), \\ V_0 = \delta_{b'} \left(\begin{array}{cc|c} e^{-b'/2} & 0 & 0 \\ 1 & e^{b'/2} & 0 \\ \hline 0 & 0 & \delta_{b'} \end{array} \right) \exp(\psi q_2),$$

(again we replaced a, b, c with b, b', b'') and $W_0 = V_0^{-1}U_0^{-1}$. The parameter κ can be expressed in terms of the third boundary parameter b'' as in eqn. (3.79):

$$(D.34) \quad \kappa = -2 \cosh \frac{b''}{2} - 2 \cosh \frac{b - b'}{2} + \psi \xi (e^{b/2} \delta_{b'} + e^{b'/2} \delta_b).$$

Let $\gamma_b, \gamma_{b'}$, and $\gamma_{b''}$ be geodesic boundaries of Λ with respective monodromies U_0, V_0 , and W_0 . In the universal cover of Λ , these geodesics lift to “unwrapped” geodesics $\widehat{\gamma}_b, \widehat{\gamma}_{b'}$, and $\widehat{\gamma}_{b''}$. The endpoints of these unwrapped geodesics can be found, by the same logic as in the bosonic case, from the eigenvectors of the monodromies acting on $u, v|\theta$. To be more precise, each monodromy has two bosonic eigenvectors, which are the endpoints of the corresponding unwrapped geodesic. (The monodromy also has a fermionic eigenvector, but it will not play a role.)

For example, one bosonic eigenvector of U_0 is $\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$. This corresponds to an endpoint of $\widehat{\gamma}_b$ at infinity. The second bosonic eigenvector is $\begin{pmatrix} -\kappa/(2 \sinh \frac{b}{2}) \\ 1 \\ \xi(1 - \delta_b e^{-b/2})^{-1} \end{pmatrix}$. So the second endpoint of $\widehat{\gamma}_b$ is

$$(D.35) \quad x_0|\theta_0 = -\kappa/(2 \sinh \frac{b}{2})|\xi(1 - \delta_b e^{-b/2})^{-1}.$$

Similarly, the bosonic eigenvectors of V_0 are $\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} -2 \sinh b'/2 \\ 1 \\ \psi(e^{b'/2} + \delta_{b'}) \end{pmatrix}$, so $\widehat{\gamma}_{b'}$ has endpoints

$$(D.36) \quad \begin{aligned} x_1|\theta_1 &= 0|0 \\ x_2|\theta_2 &= -2 \sinh(\frac{b'}{2})|\psi(e^{b'/2} + \delta_{b'}). \end{aligned}$$

Before making use of these formulas to compute the function \widehat{T} , it is convenient to first make a supersymmetry transformation $x|\theta \rightarrow x + \theta_0\theta|\theta - \theta_0$, where θ_0 was defined in eqn. (D.35). After this transformation, the endpoints of γ_b are at infinity and at

$$(D.37) \quad x'_0|\theta'_0 = -\kappa/(2 \sinh \frac{b}{2})|0.$$

Accordingly, in this description, $\widehat{\gamma}_b$ is simply a vertical straight line in $H \subset \widehat{H}$, and all of the orthogonals to $\widehat{\gamma}_b$ are likewise contained in H . Thus, we do not need to consider more general geodesics in \widehat{H} .

The supersymmetry transformation $x|\theta \rightarrow x + \theta_0\theta|\theta - \theta_0$ maps the endpoints of $\widehat{\gamma}_{b'}$ to

$$(D.38) \quad \begin{aligned} x'_1|\theta'_1 &= x_1 + \theta_0\theta_1|\theta_1 - \theta_0 \\ x'_2|\theta'_2 &= x_2 + \theta_0\theta_2|\theta_2 - \theta_0. \end{aligned}$$

The function \widehat{T} is the length of the segment of $\mathcal{B}' \subset \widehat{\gamma}_b$ consisting of points p such that the orthogonal geodesic ℓ_p to $\widehat{\gamma}_b$ at p intersects $\widehat{\gamma}_{b'}$. The endpoints of \mathcal{B}' are the points $p \in \widehat{\gamma}_b$ such that ℓ_p and $\gamma_{b'}$ satisfy the condition (D.23), which says that ℓ_p and $\widehat{\gamma}_{b'}$ just barely fail to intersect. This happens if the endpoints of ℓ_p (which are at $\theta = 0$, since ℓ_p is contained in H) are $x'_1|0$ or $x'_2|0$. Since ℓ_p is a semicircle in H centered at $x = x_0$, the same reasoning as in the bosonic case tells us that the endpoints of \mathcal{B}' are at the points in H with $(x, y) = (x'_0, x'_0 - x'_1)$ and $(x, y) = (x'_0, x'_0 - x'_2)$. Hence rather as in the bosonic case, the length of \mathcal{B}' is $\widehat{T} = \log(x'_0 - x'_2)/(x'_0 - x'_1)$. In view of eqn. (D.38), this is equivalent to $\widehat{T} = \log((x_0 - x_2 - \theta_0\theta_2)/(x_0 - x_1 - \theta_0\theta_1))$. Here in fact, $(x_0 - x_2 - \theta_0\theta_2)/(x_0 - x_1 - \theta_0\theta_1)$ is a superconformal cross ratio of the points $x_0|\theta_0$, $x_1|\theta_1$, $x_2|\theta_2$, and a fourth point at infinity. So the result is a sort of minimal supersymmetric generalization of the bosonic formula.

Using the above formulas for the endpoints and eqn. (D.34) for κ , we get

$$(D.39) \quad \begin{aligned} \widehat{T}(b, b', b''|\xi, \psi) &= \log \left(\frac{-\kappa + 4 \sinh \frac{b}{2} \sinh \frac{b'}{2} - \xi\psi(e^{b/2} + \delta_b)(e^{b'/2} + \delta_{b'})}{-\kappa} \right) \\ &= \log \frac{\cosh \frac{b''}{2} + \cosh \frac{b+b'}{2} - \frac{\xi\psi}{2} (e^{(b+b')/2} + \delta_b\delta_{b'})}{\cosh \frac{b''}{2} + \cosh \frac{b-b'}{2} + \frac{\xi\psi}{2} (e^{b/2}\delta_{b'} + e^{b'/2}\delta_b)}. \end{aligned}$$

To reduce this to the purely bosonic function $\mathcal{T}(b, b', b'')$ that appears in the final form (D.30) of the recursion relation, we have to integrate this expression over the odd moduli ξ, ψ . The torsion measure can be read off

from⁷⁷ eqn. (3.100) (where we substitute $a_t, b_t, c_t | \xi_t, \psi_t \rightarrow b, b', b'' | \xi, \psi$):

$$(D.40) \quad \tau = \frac{\delta_b \delta_{b'}}{4} e^{-(b+b')/4} (e^{b''/4} - \delta_{b''} e^{-b''/4}) [d\xi d\psi].$$

But according to eqn. (A.18), if we want to define volumes that satisfy the particularly natural normalization of eqn. (A.17), the measure we must use is not τ but

$$(D.41) \quad \mu = (2\pi)^\chi \tau = \frac{\tau}{2\pi},$$

where we use the fact that $\chi = -1$ for a three-holed sphere. So

$$(D.42) \quad \begin{aligned} \mathcal{T}(b, b', b'') &= \int_{\mathcal{M}_{b, b', b''}} d\mu \widehat{\mathbb{T}}(b, b', b'' | \xi, \psi) \\ &= -\frac{\delta_b \delta_{b'} (e^{(b+b')/4} + \delta_b \delta_{b'} e^{-(b+b')/4}) (e^{b''/4} - \delta_{b''} e^{-b''/4})}{16\pi \cosh \frac{b''}{2} + \cosh \frac{b+b'}{2}} \\ &\quad - \frac{\delta_b \delta_{b'} (e^{(b-b')/4} \delta_{b'} + e^{-(b-b')/4} \delta_b) (e^{b''/4} - \delta_{b''} e^{-b''/4})}{16\pi \cosh \frac{b''}{2} + \cosh \frac{b-b'}{2}}. \end{aligned}$$

This formula is valid for any spin structures on the boundary of Λ . For the application to the recursion relation (D.30), we want the spin structures to be all of NS type, so we take $\delta_b = \delta_{b'} = \delta_{b''} = -1$. After a few uses of the identity $2 \cosh A \cosh B = \cosh(A + B) + \cosh(A - B)$, one finds

$$(D.43) \quad \begin{aligned} \mathcal{T}(b, b', b'') &= \frac{1}{16\pi} \left(\frac{1}{\cosh \frac{b-b'+b''}{4}} + \frac{1}{\cosh \frac{b-b'-b''}{4}} \right. \\ &\quad \left. - \frac{1}{\cosh \frac{b+b'+b''}{4}} - \frac{1}{\cosh \frac{b+b'-b''}{4}} \right). \end{aligned}$$

From eqn. (D.32), we then have also

$$(D.44) \quad \begin{aligned} \mathcal{D}(b, b', b'') &= -\mathcal{T}(b, b', b'') - \mathcal{T}(b, b'', b') \\ &= -\frac{1}{8\pi} \left(\frac{1}{\cosh \frac{b'+b''-b}{4}} - \frac{1}{\cosh \frac{b'+b''+b}{4}} \right). \end{aligned}$$

⁷⁷The overall sign of this measure depends on some conventions. For example, in the symplectic approach of appendix A, the sign depends on a choice of orientation of the fermionic bundle V . Analogous issues arise in the torsion. Reversing this sign would multiply the volumes by $(-1)^\chi$, which for an orientable surface is the same as $(-1)^n$, where n is the number of boundary components.

When these results for \mathcal{D} and \mathcal{T} are inserted in the recursion relation (D.30), one finds perfect agreement with the recursion relation (5.42) derived from random matrix theory.

It is interesting to observe that the start of the topological recursion with $V_1(b) = -1/8$ can actually be computed similarly. We can build a torus with a single boundary of length b by starting with a one-holed sphere of boundaries b, b', b' and gluing together the two boundaries of length b' to make a single circle C . Repeating the derivation that leads to eqn. (D.30), assuming that spin structures are weighted with $(-1)^\zeta$, we get

$$(D.45) \quad bV_1(b) = \int_0^\infty b' db' \mathcal{D}(b, b', b').$$

Here we should divide by 2 because (due to the \mathbb{Z}_2 symmetry of the torus with one hole) we define $V_1(b)$ to be one-half of the true moduli space volume, as explained in footnote 72. But we also get a factor of 2 from the sum over spin structures. *A priori*, the spin structure on C could be of NS or R type. However, in the theory with the factor $(-1)^\zeta$, contributions in which the spin structure is of R type vanish because of the usual cancellation in the sum over the spin structure “orthogonal” to C . If the spin structure on C is of NS type; then the sum over orthogonal spin structures gives a factor of 2, compensating for the factor of 1/2 from the symmetry and leading to (D.45). One can evaluate the integral in (D.45) using eqn. (D.44), and one recovers $V_1(b) = -1/8$.

If instead one considers the theory without the factor $(-1)^\zeta$, one can recover the expectation that $V_1(b) = 0$. For this, we have to allow both types of spin structure on the circle C . The volume will then satisfy

$$(D.46) \quad b\tilde{V}_1(b) = \int_0^\infty b' db' \left(\mathcal{D}(b, b', b') - \tilde{\mathcal{D}}(b, b', b') \right),$$

where $\tilde{\mathcal{D}}(b, b', b'')$ is a function similar to $\mathcal{D}(b, b', b'')$, but for a three-holed sphere with spin structures of types NS, R, R. The minus sign has the same origin as the factor $(-1)^{w_R}$ in eqn. (3.100); it comes from the factor $-\delta$ in the formula (3.69) for the torsion of a circle. We have, as in eqn. (D.32), $\tilde{\mathcal{D}}(b, b', b'') = -\tilde{\mathcal{T}}(b, b', b'') - \tilde{\mathcal{T}}(b, b'', b')$, where now $\tilde{\mathcal{T}}(b, b', b'')$ is obtained by

setting $\delta_b = -1$, $\delta_{b'} = \delta_{b''} = 1$ in eqn. (D.42). We get

$$(D.47) \quad \tilde{\mathcal{T}}(b, b', b'') = \frac{1}{16\pi} \left(-\frac{1}{\cosh \frac{b-b'+b''}{4}} + \frac{1}{\cosh \frac{b-b'-b''}{4}} - \frac{1}{\cosh \frac{b+b'+b''}{4}} + \frac{1}{\cosh \frac{b+b'-b''}{4}} \right),$$

leading to $\tilde{\mathcal{D}}(b, b', b'') = \mathcal{D}(b, b', b'')$, and therefore to the expected $\tilde{V}_1(b) = 0$. Using the equality of \mathcal{D} and $\tilde{\mathcal{D}}$ and the vanishing of volumes in genus zero, a recursion similar to (D.30) implies that all volumes vanish in the theory without $(-1)^\zeta$.

Appendix E. Comments on nonperturbative effects

In the main text of the paper, we focused on aspects of JT gravity and of matrix integrals that are perturbative in e^{-S_0} . In this appendix, we make a few comments on nonperturbative effects. For the most part, we will consider an Altland-Zirnbauer-type matrix integral, with measure of the form

$$(E.1) \quad \prod_{i < j} |\lambda_i - \lambda_j|^\beta \prod_i \lambda_i^{\frac{\alpha-1}{2}} e^{-LV(\lambda_i)} d\lambda_i, \quad \lambda_i > 0,$$

and with $\alpha = 0, 1, 2$ and $\beta = 2$. These three cases are dual to JT supergravity theories without a crosscap divergence. We will also comment on a fourth case with this property in section E.3.

E.1. The Bessel curve

A useful starting point is the solvable “Bessel” spectral curve $y^2 = -1/x$, or equivalently $\rho_0 = 1/(\pi\sqrt{x})$. This spectral curve describes the small x region of any Altland-Zirnbauer ensemble, including the ones dual to JT gravity. It plays the role of the “Airy” curve $y^2 = -x$ in the standard Dyson ensembles. A matrix integral with spectral curve $y^2 = -1/x$ can be obtained as a double-scaled limit of a conventional matrix integral with $V(\lambda) = \lambda$. For this potential, the orthogonal polynomials are the $\tilde{\alpha}$ -Laguerre polynomials, with $\tilde{\alpha} = (\alpha - 1)/2$. Using asymptotics for these functions, Nagao and Slevin

[110] computed the exact density of eigenvalues in the double-scaled limit:

$$(E.2) \quad \langle \rho(x) \rangle = e^{2S_0} \left(J_{\tilde{\alpha}+1}(\xi)^2 - J_{\tilde{\alpha}}(\xi)J_{\tilde{\alpha}+2}(\xi) + 2 \frac{J_{\tilde{\alpha}}(\xi)J_{\tilde{\alpha}+1}(\xi)}{\xi} \right)$$

$$\xi = 2e^{S_0}\sqrt{x}.$$

This function has an asymptotic expansion for large e^{S_0} . The expansion contains two types of terms: odd powers of e^{-S_0} , and both even and odd powers of e^{-S_0} multiplied by a nonperturbative term oscillating with fixed frequency. The leading terms of both types are

$$(E.3) \quad \langle \rho(x) \rangle \approx \frac{e^{S_0}}{\pi\sqrt{x}} + \dots + \frac{\sin\left(-\frac{\pi}{2}\alpha + 4e^{S_0}\sqrt{x}\right)}{4\pi x} + \dots$$

The terms that form the first class of dots can be obtained from the recursive treatment of the loop equations that we described in section 4. The oscillating terms are nonperturbative in e^{-S_0} and are not obtained this way.

When $\alpha \in \{0, 2\}$, we argued in section 5.3.3 that after double-scaling, the only nonzero terms in the genus expansion were the disk $R_0(x)$, the cylinder $R_0(x_1, x_2)$, and the crosscap $R_{\frac{1}{2}}(x)$. Indeed, for $\alpha \in \{0, 2\}$, the dots in (E.3) vanish and the two terms are the exact answer. We can use this to compute the resolvent $R(x) = \int_0^\infty dx' \frac{\rho(x')}{x-x'}$. It is natural to define the resolvent with a branch cut along the positive real axis, where the eigenvalue density is nonzero. Then

$$(E.4) \quad \langle R(x) \rangle = -\frac{e^{S_0}}{\sqrt{-x}} \pm \frac{1}{4x} \mp \frac{\exp(-4e^{S_0}\sqrt{-x})}{4x}.$$

where the top sign is for $\alpha = 0$ and the bottom sign is for $\alpha = 2$. The first term is $e^{S_0}y(x)$. The second term is the crosscap contribution $R_{\frac{1}{2}}(x)$, and the final term is a nonperturbative correction. The truncation of the perturbative series is precisely consistent with the expectations of section 5.3.3.

When $\alpha = 1$, the Bessel functions in (E.2) have a nontrivial asymptotic expansion, which implies an expansion of the resolvent that we will find useful below. To compute this, it is convenient to take a slight detour and first compute the expectation value of $Z(\beta) = \int_0^\infty dx \rho(x')e^{-\beta x'}$. This can be done by inserting an integral representation

$$(E.5) \quad J_\nu(\xi) = \frac{(\frac{1}{2}\xi)^\nu}{2\pi i} \int_{\epsilon+i\mathbb{R}} \exp\left(t - \frac{\xi^2}{4t}\right) \frac{dt}{t^{\nu+1}}$$

of each of the Bessel functions in (E.2), with integration parameters t_1, t_2 for the two factors of J . Then $\langle Z(\beta) \rangle$ becomes an integral over x, t_1, t_2 . The x and t_1 integrals can be done easily. After a change of variables, the final t_2 integral leads to

$$(E.6) \quad \langle Z(\beta) \rangle = \int_0^a \frac{du}{2\pi} \sqrt{\frac{a-u}{u}} e^{-u}, \quad a = \frac{4e^{2S_0}}{\beta}.$$

To find the e^{-S_0} expansion of this formula, we expand the square root in a power series in u/a , and integrate from zero to infinity. At each order in $1/a$, this gives a simple expression involving gamma functions. Using $R(x) = -\int_0^\infty d\beta e^{\beta x} Z(\beta)$, and integrating the expansion of $\langle Z(\beta) \rangle$ term-by-term, one finds a result that can be written (for $g \geq 1$) as

$$(E.7) \quad R_g(x) = \frac{(2g)!}{2^{8g}(2g-1)} \binom{2g}{g}^2 \frac{1}{(-x)^{g+\frac{1}{2}}}.$$

E.2. More general spectral curves

Still with $\beta = 2$ and $\alpha = 0, 1, 2$, we would like to generalize (E.3) beyond the curve $y^2 = -1/x$. It is convenient to start out by considering the operator $\psi(x) = e^{-LV(x)/2} \det(x - H)$. For the Bessel curve $y^2 = -1/x$, the expectation value of $\psi(x)$ is

$$(E.8) \quad \langle \psi(x) \rangle = (\text{const.}) \frac{J_{\tilde{\alpha}}(\xi)}{\xi^{\tilde{\alpha}}} \approx (\text{const.}) \left[\frac{\cos(2e^{S_0}\sqrt{x} - \frac{\pi}{4}\alpha)}{x^{\frac{\alpha}{4}}} + \dots \right]$$

where ξ is defined in (E.2), and the final expression is the leading asymptotics for large e^{S_0} and positive x .⁷⁸ For a general spectral curve, we can't hope to find an exact expression, but we would like to generalize the leading asymptotics. To do so, one can write

$$(E.9) \quad \det(x - H) = \exp\left(\int^x dx' R(x')\right)$$

⁷⁸For a conventional finite L matrix integral of type $(\alpha, 2)$, the expectation value of $\det(x - H)$ is equal to the order L monic polynomial from the set of polynomials orthogonal with respect to the measure $x^{(\alpha-1)/2} e^{-LV(x)}$ on the positive real axis. For the potential $V(x) = x$, these polynomials are the $\tilde{\alpha}$ -Laguerre polynomials. Double-scaling to obtain $y^2 = -1/x$ gives the Bessel function quoted.

in terms of the resolvent R , and then expand in correlation functions of the resolvent using

$$(E.10) \quad \langle e^X \rangle = \exp \left(\langle X \rangle + \frac{1}{2} \langle X^2 \rangle_c + \dots \right).$$

In this expansion, terms of order e^{S_0} and order one come from only three places: $R_0(x)$, $R_{\frac{1}{2}}(x)$, and $R_0(x_1, x_2)$. So to compute the expectation value of the determinant to this order, we exponentiate these terms. This leads to an approximate expression for $\langle \psi(x) \rangle$ in terms of the function $\Psi(x)$, defined by

$$(E.11) \quad \begin{aligned} \Psi(x) &= \exp \left[\text{Disk}(x) + \text{Xcap}(x) + \frac{1}{2} \text{Cyl}(x, x) \right] \\ &= (\text{const.}) \frac{\exp(\text{Disk}(x))}{(-x)^{\frac{\alpha}{4}}}. \end{aligned}$$

Here, the disk, crosscap and cylinder functions are defined for $x < 0$ by

$$(E.12) \quad \text{Disk}(x) = \int_0^x dx' \left(e^{S_0} R_0(x') - \frac{LV'(x')}{2} \right) = e^{S_0} \int_0^x dx' y(x')$$

$$(E.13) \quad \text{Xcap}(x) = \int^x dx' R_{\frac{1}{2}}(x) = -\frac{\alpha-1}{4} \log(-x) + (\text{const.})$$

$$(E.14) \quad \begin{aligned} \text{Cyl}(x_1, x_2) &= \int_{-\infty}^{x_1} dx'_1 \int_{-\infty}^{x_2} dx'_2 R_0(x'_1, x'_2) \\ &= -\log(\sqrt{-x_1} + \sqrt{-x_2}) + (\text{const.}). \end{aligned}$$

The main complication is that that $\langle \psi(x) \rangle$ is entire, while the function $\Psi(x)$ is multi-valued. As explained in [111], one has to sum over branch choices for $\Psi(x)$ with the right coefficients. These can be determined by matching to the formula (E.8) for small x . A shortcut for doing this is to start with (E.8) and simply replace the factor $\pm 2ie^{S_0} \sqrt{x}$ by its generalization $\text{Disk}(x) = \pm i \int_0^x dx' \rho_0(x')$. This leads to the following formula for the leading asymptotics for large e^{S_0} and positive x :

$$(E.15) \quad \langle \psi(x) \rangle \approx (\text{const.}) \frac{\cos \left(-\frac{\pi}{4} \alpha + \pi e^{S_0} \int_0^x dx \rho_0(x) \right)}{x^{\frac{\alpha}{4}}} + \dots$$

This simple replacement is possible because the crosscap and cylinder terms $R_{\frac{1}{2}}(x)$ and $R_0(x_1, x_2)$ are completely universal for double-scaled $(\alpha, 2)$ theories. So, to this order, the only dependence on the spectral curve comes from

the disk amplitude, which is easy to isolate as the term in the exponential multiplying e^{S_0} .

So far we have just discussed the determinant operator, but the logic for the density of eigenvalues is similar. One can write $\langle \rho(x) \rangle$ in terms of correlation functions of determinants, so we expect the same universality to apply. Starting with the oscillating term in (E.3) and making the replacement of $\pm 2ie^{S_0}\sqrt{x}$ by $\text{Disk}(x) = \pm i \int_0^x dx' \rho_0(x')$, one finds

$$(E.16) \quad \langle \rho(x) \rangle \approx e^{S_0} \rho_0(x) + \dots + \frac{\sin\left(-\frac{\pi}{2}\alpha + 2\pi e^{S_0} \int_0^x dx' \rho_0(x')\right)}{4\pi x} + \dots$$

We emphasize that this formula is only supposed to apply for double-scaled $(\alpha, 2)$ theories.

If we further specialize to the cases $\alpha \in \{0, 2\}$, both sets of dots should vanish in perturbation theory. We don't want to claim that (E.16) is the exact answer though, because there could be "instanton" corrections related to nontrivial critical points of the effective potential of the matrix integral. However, unlike the oscillating term in (E.16), these instanton effects (if present) are expected to be exponentially small, of order $\exp(-\#e^{S_0})$.

E.3. A case without the oscillating term

There is one more JT supergravity theory without a crosscap divergence: the case without time-reversal symmetry, and with equal weighting for even and odd spin structures in the bulk. It makes sense to ask about nonperturbative effects for this model. The random matrix ensemble for the supercharge Q is a standard Dyson $\beta = 2$ ensemble, double-scaled in such a way that both endpoints of the cut move to $\pm\infty$. As we saw in section 5.2.1, all perturbative correlators vanish except $R_0(x)$ and $R_0(x_1, x_2)$.

We expect the oscillating contribution in (E.16) to be absent for this case. The reason is that for a standard $\beta = 2$ ensemble, the oscillating corrections to the density of eigenvalues fade as we move farther from the endpoints of the cut, and in a limit where the endpoints have been scaled away, the oscillating correction should vanish. Note, however, that in $\langle \rho(x) \rangle$ there could still be $\exp(-\#e^{S_0})$ instanton effects that arise from nontrivial critical points of the effective potential.

E.4. Comments on volumes

The formula (E.7) can be used to determine the large b asymptotics of volumes $V_g(b)$ for the case that corresponds to a matrix integral with $(\alpha, \beta) = (1, 2)$. This is the case discussed in section 5.2.2: orientable surfaces and a bulk topological field theory $(-1)^\zeta$.

The logic is as follows. First, since the spectral curve for JT gravity reduces for small x to a multiple of the Bessel curve $y^2 = -1/x$, the small x asymptotics of the resolvent are determined by the formula (E.7):

$$(E.17) \quad R_{g, \text{SJT}}(x) = 2 \cdot \sqrt{2}^{2g-1} \cdot R_{g, \text{Bessel}}(x) + \dots$$

$$(E.18) \quad = \frac{\sqrt{2}}{2^{7g}} \frac{(2g)!}{2g-1} \binom{2g}{g}^2 \frac{1}{(-x)^{g+\frac{1}{2}}} + \dots$$

where the dots are subleading by powers of x as $x \rightarrow 0$. In the first line, the first factor of two is to account for the difference between $R_{g, \text{SJT}}$ and $R_{g, \text{M}}$ as in (5.33). The second factor $\sqrt{2}^{2g-1}$ is to account for the fact that for small x , the relevant spectral curve (5.32) limits to $y^2 = -1/(2x)$ instead of $y^2 = -1/x$. To convert we need to rescale y by $\sqrt{2}$, which introduces the factor shown. In the second line we inserted (E.7) for $R_{g, \text{Bessel}}(x)$.

Now, $R_{g, \text{SJT}}(x)$ is an integral transform of $V_g(b)$, and in order to produce the small x limit (E.18), we need the leading term in $V_g(b)$ as $b \rightarrow \infty$ to be

$$(E.19) \quad V_g(b) = -\frac{2^{2-7g}g}{2g-1} \binom{2g}{g}^2 b^{2g-2} + \dots$$

where the dots are subleading by powers of $1/b^2$ as $b \rightarrow \infty$. We emphasize that this formula is valid for arbitrary g , but it only gives the leading term in powers of b .

In the bosonic case, an analogous formula for the leading coefficient in powers of b is also known exactly. Apart from this formula, and low-genus results, little is known exactly. However, there are good large genus approximate formulas, based on Zograf's [112] conjectures for the asymptotics of intersection numbers. These imply asymptotic formulas for the bosonic volumes $V_g(b_1, \dots, b_n)$ for large g , with the restriction that all lengths b_i be much smaller than g . Zograf's formulas were motivated by comparison to explicit low-genus computations, but aspects of his conjectures were later established rigorously [113, 114]. For the special case of one boundary, a formula was proposed in [35] for $V_g(b)$ in a somewhat larger domain: large g

and arbitrary b . This formula was inspired by a matrix integral calculation and has not been established rigorously.

For the volumes of super moduli space $V_g(b)$, we used a method similar to that of [35], together with (E.16), to generate the following conjecture for large g and arbitrary b :

$$(E.20) \quad V_g(b) \approx -\frac{\Gamma(2g-1)}{2^g \pi^{3-2g} i} \oint \frac{dz}{bz} \frac{\sinh(bz)}{(\sin(2\pi z))^{2g-1}}, \quad g \gg 1.$$

The contour integral around the origin reduces to a residue, which is an even polynomial in b . The method that led us to (E.20) is not rigorous, and we prefer not to try to present a partial derivation. But we feel that it is worth writing the formula, because agreement with low genus exact results (computed by topological recursion) seems promising. As a function of b , the maximum percentage error for $g = 12$ is around 2.1%. One can also try to extrapolate the error in $1/g$, and we found the best results by holding bg fixed in the extrapolation. Using a five-term extrapolation in $1/g$ from $g = 7, \dots, 12$, the maximum over b of the extrapolated error is around 0.015%.

As another piece of evidence, one can check that the leading term in (E.20) for large b is proportional to b^{2g-2} , and the coefficient agrees with the exact result (E.19) for large g . This is encouraging, especially since the large b region is where the largest percentage error is for $g = 1, \dots, 12$. Finally, as an opposite limit one can take g to be much larger than b . This is analogous to the limit where the Zograf conjecture applies in the ordinary bosonic case. In this limit, saddle points at $z = \pm \frac{1}{4}$ dominate the integral in (E.20), and we find

$$(E.21) \quad V_g(b) \approx -\frac{\Gamma(2g - \frac{3}{2})}{2^{g-\frac{5}{2}} \pi^{\frac{7}{2}-2g}} \frac{\sinh(\frac{b}{4})}{b} \quad g \gg b, \quad g \gg 1.$$

We don't have a known result to compare to in this case, but we note that the dependence on b is similar to the bosonic case, although with $\sinh(b/4)$ instead of the bosonic $\sinh(b/2)$.

Appendix F. Comments on the minimal string

A matrix integral is described by a discrete choice of symmetry class, together with a choice of spectral curve. In the main text of this paper, we encountered all ten of the standard symmetry classes of matrix integral, but

we only studied two essentially different spectral curves: one for the bosonic models and one for the supersymmetric ones.

To access different spectral curves, one would have to replace JT gravity or supergravity with a different bulk theory. A class of such bulk theories has been studied extensively in the literature: the “minimal string” theories. These are 2d worldsheet gravity theories that consist of a $(2, p)$ minimal model, together with Liouville theory and the reparametrization ghosts. There are both bosonic and supersymmetric versions of these theories. We will briefly comment on both cases.

In the bosonic case, the connection was made in [62, 64] between the $\beta = 1, 4$ ensembles and unoriented versions of the bulk minimal string theories, with the $\beta = 4$ case distinguished by including a $(-1)^{n_c}$ factor. In the context of the present paper, there is one interesting detail. Up to arbitrary choices of normalization, the leading density of eigenvalues for the bosonic $(2, p)$ minimal string is [115, 116]

$$(F.1) \quad \rho_0(x) = \frac{1}{4\pi^2} \sinh \left(\frac{p}{2} \operatorname{arccosh} \left(1 + \frac{8\pi^2}{p^2} x \right) \right)$$

where p is an odd integer. In the limit $p \rightarrow \infty$, this reduces to the spectral curve of JT gravity [35]. However, for fixed p , the large x behavior is dramatically modified. Indeed, substituting (F.1) and $y = \pm i\pi\rho_0$ into the formula for the “crosscap” contribution (5.16), we find a finite integral. As $p \rightarrow \infty$, one has

$$(F.2) \quad R_{\frac{1}{2}}(x) = \frac{\log(p)}{\sqrt{-x}} + (\text{finite})$$

so crosscap divergence in the JT theory is regulated in the finite p minimal string.

Now we turn to the supersymmetric case. For orientable surfaces, two versions of the supersymmetric minimal string have been defined. These are referred to as 0A and 0B, and they differ in the weighting of the sum over spin structures. The 0B theory involves a sum with uniform weighting, and the 0A theory includes a factor of $(-1)^\zeta$.⁷⁹

⁷⁹In general for type 0 string theories (unlike in type II) the left-moving and right-moving worldsheet fermions experience the same spin structure, and the GSO projection depends only on the total (left + right) fermion number; see section 10.6 of [117]. The projection simply requires that NS sector states must be bosonic, and R sector states must be fermionic in type 0A and bosonic in type 0B. This projection is equivalent to summing over all spin structures, with equal weighting

These two versions of the minimal string were related to matrix integrals in [116, 118–120], building in part on earlier work [98–100, 121, 122]. The type 0B theory (with an appropriate sign of the super-Liouville cosmological constant μ) was related to a double-scaled unitary matrix integral with no endpoints in the eigenvalue distribution. In the double-scaled limit, the distinction between unitary and hermitian matrix integrals disappears, so this is the same type of matrix ensemble that we found for the supercharge Q in the case without $(-1)^\zeta$. The type 0A theory was related to a double-scaled complex matrix integral. For square complex matrices, this is the same thing as an $(\alpha, \beta) = (1, 2)$ Altland-Zirnbauer ensemble, which is the symmetry class that we found in the case with $(-1)^\zeta$. For rectangular matrices, this is the generalization that we studied in section 5.5.

So the symmetry classes that we identified in this paper for the supersymmetric case without time-reversal symmetry are consistent with the previous results. We expect that large p limits of these minimal string theories (with the appropriate sign of μ) will coincide with JT supergravity.

The supersymmetric minimal string has a time-reversal symmetry, and one can gauge this symmetry, leading to a theory that includes a sum over unorientable as well as orientable surfaces. In this situation, each of the 0A and 0B theories have a four-fold bifurcation, for a total of eight different theories. These differ by weighting the pin^- sum by a factor $\exp(-i\pi\eta N'/2)$, where N' is an integer mod eight. The cases with odd N' generalize the 0A theory, and the cases with even N' generalize 0B. We expect these to be related to matrix integrals of the classes described in table 4.

References

- [1] C. Teitelboim, *Gravitation and Hamiltonian structure in two space-time dimensions*, Phys. Lett. **B126** (1983), 41–45.

for type 0B, and with weighting $(-1)^\zeta$ for type 0A. Note that the GSO projection that defines type II string theories can also be understood as a particular sum over spin structures. However, one has to define separate spin structures for the left-moving and right-moving worldsheet fermions. Corresponding super Riemann surfaces (and their moduli spaces) can be defined, with no relationship between holomorphic and antiholomorphic odd coordinates. Such “cs supermanifolds” are essentially different from the real supermanifolds that arise in JT supergravity, and in 0A and 0B theory, on an orientable manifold. (Some of the subtleties familiar in superstring theory do become relevant in the unorientable case, because pin^- structures in two dimensions are not real.)

- [2] R. Jackiw, *Lower dimensional gravity*, Nucl. Phys. **B252** (1985), 343–356.
- [3] A. Almheiri and J. Polchinski, *Models of AdS_2 backreaction and holography*, JHEP **11** (2015) 014.
- [4] K. Jensen, *Chaos in AdS_2 holography*, Phys. Rev. Lett. **117** (2016), no. 11, 111601.
- [5] J. Maldacena, D. Stanford, and Z. Yang, *Conformal symmetry and its breaking in two dimensional Nearly Anti-de-Sitter space*, PTEP **2016** (2016), no. 12, 12C104.
- [6] J. Engelsöy, T. G. Mertens, and H. Verlinde, *An investigation of AdS_2 backreaction and holography*, JHEP **07** (2016) 139.
- [7] S. Sachdev and J.-w. Ye, *Gapless spin fluid ground state in a random, quantum Heisenberg magnet*, Phys. Rev. Lett. **70** (1993) 3339.
- [8] A. Kitaev, *A simple model of quantum holography*, talks at KITP, April 7, 2015 and May 27, 2015.
<http://online.kitp.ucsb.edu/online/entangled15/kitaev/>
<http://online.kitp.ucsb.edu/online/entangled15/kitaev2/>
- [9] A. Kitaev and S. J. Suh, *The soft mode in the Sachdev-Ye-Kitaev model and its gravity dual*, JHEP **05** (2018) 183.
- [10] J. Maldacena and D. Stanford, *Remarks on the Sachdev-Ye-Kitaev model*, Phys. Rev. **D94** (2016), no. 10, 106002.
- [11] G. Sárosi, *AdS_2 holography and the SYK model*, PoS **Modave2017** (2018) 001.
- [12] D. Bagrets, A. Altland, and A. Kamenev, *Sachdev-Ye-Kitaev model as Liouville quantum mechanics*, Nucl. Phys. **B911** (2016), 191–205.
- [13] J. S. Cotler, G. Gur-Ari, M. Hanada, J. Polchinski, P. Saad, S. H. Shenker, D. Stanford, A. Streicher, and M. Tezuka, *Black holes and random matrices*, JHEP **05** (2017) 118.
- [14] D. Bagrets, A. Altland, and A. Kamenev, *Power-law out of time order correlation functions in the SYK model*, Nucl. Phys. **B921** (2017), 727–752.
- [15] D. Stanford and E. Witten, *Fermionic localization of the Schwarzian theory*, JHEP **10** (2017) 008.

- [16] V. V. Belokurov and E. T. Shavgulidze, *Exact solution of the Schwarzian theory*, Phys. Rev. **D96** (2017), no. 10, 101701.
- [17] T. G. Mertens, G. J. Turiaci, and H. L. Verlinde, *Solving the Schwarzian via the conformal bootstrap*, JHEP **08** (2017) 136.
- [18] A. Kitaev and S. J. Suh, *Statistical mechanics of a two-dimensional black hole*, arXiv:1808.07032 [hep-th].
- [19] Z. Yang, *The quantum gravity dynamics of near extremal black holes*, arXiv:1809.08647 [hep-th].
- [20] L. V. Iliesiu, S. S. Pufu, H. Verlinde, and Y. Wang, *An exact quantization of Jackiw-Teitelboim gravity*, arXiv:1905.02726 [hep-th].
- [21] D. Harlow and D. Jafferis, *The factorization problem in Jackiw-Teitelboim gravity*, arXiv:1804.01081 [hep-th].
- [22] J. Lin, *Entanglement entropy in Jackiw-Teitelboim gravity*, arXiv:1807.06575 [hep-th].
- [23] A. Blommaert, T. G. Mertens, and H. Verschelde, *Fine structure of Jackiw-Teitelboim quantum gravity*, arXiv:1812.00918 [hep-th].
- [24] T. G. Mertens, *The Schwarzian theory – origins*, JHEP **05** (2018) 036.
- [25] A. Blommaert, T. G. Mertens, and H. Verschelde, *The Schwarzian theory — A Wilson line perspective*, JHEP **12** (2018) 022. arXiv:1806.07765 [hep-th].
- [26] A. Blommaert, T. G. Mertens, and H. Verschelde, *Clocks and Rods in Jackiw-Teitelboim quantum gravity*, arXiv:1902.11194 [hep-th].
- [27] H. W. Lin, J. Maldacena, and Y. Zhao, *Symmetries near the horizon*, arXiv:1904.12820 [hep-th].
- [28] A. R. Brown, H. Gharibyan, H. W. Lin, L. Susskind, L. Thorlacius, and Y. Zhao, *Complexity of Jackiw-Teitelboim gravity*, Phys. Rev. **D99** (2019), no. 4, 046016.
- [29] H. A. González, D. Grumiller, and J. Salzer, *Towards a bulk description of higher spin SYK*, JHEP **05** (2018) 083.
- [30] S. Forste and I. Golla, *Nearly AdS₂ sugra and the super-Schwarzian*, Phys. Lett. **B771** (2017), 157–161.
- [31] J. Maldacena and X.-L. Qi, *Eternal traversable wormhole*, arXiv:1804.00491 [hep-th].

- [32] P. Saad, S. H. Shenker, and D. Stanford, *A semiclassical ramp in SYK and in gravity*, arXiv:1806.06840 [hep-th].
- [33] J. Maldacena, G. J. Turiaci, and Z. Yang, *Two dimensional Nearly de Sitter gravity*, arXiv:1904.01911 [hep-th].
- [34] J. Cotler, K. Jensen, and A. Maloney, *Low-dimensional de Sitter quantum gravity*, arXiv:1905.03780 [hep-th].
- [35] P. Saad, S. H. Shenker, and D. Stanford, *JT gravity as a matrix integral*, arXiv:1903.11115 [hep-th].
- [36] B. Eynard and N. Orantin, *Weil-Petersson volume of moduli spaces, Mirzakhani's recursion and matrix models*, arXiv:0705.3600 [math-ph].
- [37] M. Mirzakhani, *Simple geodesics and Weil-Petersson volumes of moduli spaces of bordered riemann surfaces*, *Inventiones mathematicae* **167** (2007), no. 1, 179–222.
- [38] B. Eynard and N. Orantin, *Invariants of algebraic curves and topological expansion*, arXiv:math-ph/0702045.
- [39] B. Eynard, *All genus correlation functions for the hermitian 1-matrix model*, arXiv:hep-th/0407261.
- [40] D. Montano, K. Aoki, and J. Sonnenschein, *Topological supergravity in two-dimensions*, *Phys. Lett.* **B247** (1990), 64–70.
- [41] A. H. Chamseddine, *Superstrings in arbitrary dimensions*, *Phys. Lett.* **B258** (1991), 97–103.
- [42] D. Cangemi and M. Leblanc, *Two-dimensional gauge theoretic supergravities*, *Nucl. Phys.* **B420** (1994), 363–378.
- [43] F. J. Dyson, *Statistical theory of the energy levels of complex systems. I*, *J. Math. Phys.* **3** (1962), 140–156.
- [44] A. Altland and M. R. Zirnbauer, *Nonstandard symmetry classes in mesoscopic normal-superconducting hybrid structures*, *Physical Review B* **55** (1997), no. 2, 1142.
- [45] J. J. M. Verbaarschot and I. Zahed, *Spectral density of the QCD Dirac operator near zero virtuality*, *Phys. Rev. Lett.* **70** (1993), 3852–3855.
- [46] J. J. M. Verbaarschot, *The Spectrum of the QCD Dirac operator and chiral random matrix theory: The Threefold way*, *Phys. Rev. Lett.* **72** (1994), 2531–2533.

- [47] A. A. Migdal, *Loop equations and $1/n$ expansion*, Physics Reports **102** (1983), no. 4, 199–290.
- [48] M. F. Atiyah, V. K. Patodi, and I. M. Singer, *Spectral asymmetry and Riemannian Geometry 1*, Math. Proc. Cambridge Phil. Soc. **77** (1975) 43.
- [49] T. Fukuyama and K. Kamimura, *Gauge theory of two-dimensional gravity*, Phys. Lett. **160B** (1985), 259–262.
- [50] A. H. Chamseddine and D. Wyler, *Gauge theory of topological gravity in $(1 + 1)$ -dimensions*, Phys. Lett. **B228** (1989), 75–78.
- [51] A. S. Schwarz, *The partition function of degenerate quadratic functional and Ray-Singer invariants*, Lett. Math. Phys. **2** (1978), 247–252.
- [52] P. Norbury, *Lengths of geodesics on non-orientable hyperbolic surfaces*, Geometriae Dedicata **134** (2008), no. 1, 153–176.
- [53] L. Chekhov and B. Eynard, *Matrix eigenvalue model: Feynman graph technique for all genera*, JHEP **12** (2006) 026.
- [54] L. O. Chekhov, B. Eynard, and O. Marchal, *Topological expansion of β -ensemble model and quantum algebraic geometry in the sectorwise approach*, Theor. Math. Phys. **166** (2011), 141–185.
- [55] O. Marchal, *One-cut solution of the β ensembles in the Zhukovsky variable*, J. Stat. Mech. **1201** (2012) P01011.
- [56] P. Norbury, *A new cohomology class on the moduli space of curves*, arXiv:1712.03662, (2017).
- [57] N. Do and P. Norbury, *Topological recursion on the Bessel curve*, Commun. Num. Theor. Phys. **12** (2018), 53–73.
- [58] N. Do and P. Norbury, *Topological recursion for irregular spectral curves*, arXiv:1412.8334 [math-GT].
- [59] Y. Huang, R. C. Penner, and A. Zeitlin, *McShane identity for 1-cusped super tori*, to appear.
- [60] M. Mulase and A. Waldron, *Duality of orthogonal and symplectic matrix integrals and quaternionic feynman graphs*, Communications in mathematical physics **240** (2003), no. 3, 553–586.

- [61] M. R. Zirnbauer, *Riemannian symmetric superspaces and their origin in random-matrix theory*, J. Math. Phys. **37** (1996), 4986–5018.
- [62] G. R. Harris and E. J. Martinec, *Unoriented strings and matrix ensembles*, Phys. Lett. **B245** (1990), 384–392.
- [63] E. Brezin and H. Neuberger, *Large N scaling limits of symmetric matrix models as systems of fluctuating unoriented surfaces*, Phys. Rev. Lett. **65** (1990), 2098–2101.
- [64] E. Brezin and H. Neuberger, *Multicritical points of unoriented random surfaces*, Nucl. Phys. **B350** (1991), 513–553.
- [65] Y.-Z. You, A. W. W. Ludwig, and C. Xu, *Sachdev-Ye-Kitaev model and thermalization on the boundary of many-body localized fermionic symmetry protected topological states*, Phys. Rev. **B95** (2017), no. 11, 115150.
- [66] T. Li, J. Liu, Y. Xin, and Y. Zhou, *Supersymmetric SYK model and random matrix theory*, JHEP **06** (2017) 111.
- [67] T. Kanazawa and T. Wettig, *Complete random matrix classification of SYK models with $\mathcal{N} = 0, 1$ and 2 supersymmetry*, JHEP **09** (2017) 050. [arXiv:1706.03044 \[hep-th\]](#).
- [68] F. Sun and J. Ye, *Periodic Table of SYK and supersymmetric SYK*, [arXiv:1905.07694 \[cond-mat.str-el\]](#).
- [69] A. Y. Kitaev, *Unpaired Majorana fermions in quantum wires*, Phys. Usp. **44** (2001), no. 10S, 131–136.
- [70] A. Kapustin and N. Seiberg, *Coupling a QFT to a TQFT and duality*, JHEP **04** (2014) 001.
- [71] R. Pandharipande, J. P. Solomon, and R. J. Tessler, *Intersection theory on moduli of disks, open KdV and Virasoro*, [arXiv:1409.2191 \[math.SG\]](#).
- [72] R. Dijkgraaf and E. Witten, *Developments in topological gravity*, [arXiv:1804.03275 \[hep-th\]](#).
- [73] L. Fidkowski and A. Kitaev, *The effects of interactions on the topological classification of free fermion systems*, Phys. Rev. **B81** (2010) 134509.
- [74] A. Prakash, J. Wang, and T.-C. Wei, *Unwinding short-range entanglement*, Phys. Rev. **B98** (2018), no. 12, 125108.

- [75] M. F. Atiyah, R. Bott, and A. Shapiro, *Clifford modules*, *Topology* **3** (1964), S3–S38.
- [76] E. Witten, *Fermion path integrals and topological phases*, *Rev. Mod. Phys.* **88** (2016), no. 3, 035001.
- [77] G. 't Hooft, *A planar diagram theory for strong interactions*, *Nucl. Phys.* **B72** (1974) 461.
- [78] E. Brezin, C. Itzykson, G. Parisi, and J. B. Zuber, *Planar Diagrams*, *Commun. Math. Phys.* **59** (1978) 35.
- [79] L. Fidkowski and A. Kitaev, *The effects of interactions on the topological classification of free fermion systems*, *Phys. Rev.* **B81** (2010) 134509.
- [80] A. Kapustin, R. Thorngren, A. Turzillo, and Z. Wang, *Fermionic symmetry protected topological phases and cobordisms*, *JHEP* **12** (2015) 052.
- [81] W. Fu, D. Gaiotto, J. Maldacena, and S. Sachdev, *Supersymmetric Sachdev-Ye-Kitaev models*, *Phys. Rev.* **D95** (2017), no. 2, 026009. [Addendum: *Phys. Rev.* **D95** (2017), no. 6, 069904].
- [82] K. Reidemeister, *Homotopieringe und linsenräume*, *Hamburger Abhandl.* **11** (Dec, 1935), no. 1, 102–109.
- [83] D. B. Ray and I. M. Singer, *R-torsion and the Laplacian on Riemannian manifolds*, *Advances in Mathematics* **7** (1971), no. 2, 145–210.
- [84] E. Witten, *Quantum field theory and the Jones polynomial*, *Commun. Math. Phys.* **121** (1989), 351–399.
- [85] J. Cheeger, *Analytic torsion and the heat equation*, *Annals of Mathematics* **109** (1979), no. 2, 259–321.
- [86] W. Müller, *Analytic torsion and R-torsion of riemannian manifolds*, *Advances in Mathematics* **28** (1978), no. 3, 233–305.
- [87] W. Müller, *Analytic torsion and R-torsion for unimodular representations*, *Journal of the American Mathematical Society* **6** (1993), no. 3, 721–753.
- [88] J.-M. Bismut and W. Zhang, *Métriques de Reidemeister et métriques de Ray-Singer sur le déterminant de la cohomologie d'un fibré plat: une extension d'un résultat de Cheeger et Müller*, *Comptes rendus de l'Académie des sciences. Série 1, Mathématique* **313** (1991), no. 11, 775–782.

- [89] J.-M. Bismut and W. Zhang, *An extension of a theorem by Cheeger and Müller*, Astérisque **205** (1992) 235.
- [90] E. Witten, *On quantum gauge theories in two-dimensions*, Commun. Math. Phys. **141** (1991), 153–209.
- [91] A. A. Migdal, *Gauge transitions in gauge and spin lattice systems*, Sov. Phys. JETP **42** (1975) 743. [Zh. Eksp. Teor. Fiza. **69** (1975) 1457].
- [92] J. Brüning and X. Ma, *On the gluing formula for the analytic torsion*, Mathematische Zeitschrift **273** (2013), no. 3-4, 1085–1117.
- [93] M. Gendulphe, *What’s wrong with the growth of simple closed geodesics on nonorientable hyperbolic surfaces*, arXiv:1706.08798, (2017).
- [94] P. Deligne and J. W. Morgan, *Notes on supersymmetry (following Joseph Bernstein)*, Quantum Fields and Strings: A Course for Mathematicians **1** (1999) 2.
- [95] F. F. T. Voronov, *Quantization on supermanifolds and an analytic proof of the Atiyah-Singer index theorem*, J. Soviet Math. **186** (1994), 3–118.
- [96] F. F. T. Voronov and A. V. Zorich, *Integral transformations of pseudodifferential forms*, Russian Math. Surveys **41** (1986), 221–2.
- [97] R. C. Penner and A. M. Zeitlin, *Decorated super-Teichmüller space*, J. Diff. Geom. **111** (2019), no. 3, 527–566.
- [98] T. R. Morris, *Checked surfaces and complex matrices*, Nucl. Phys. **B356** (1991), 703–728.
- [99] S. Dalley, C. V. Johnson, and T. R. Morris, *Multicritical complex matrix models and nonperturbative 2-D quantum gravity*, Nucl. Phys. **B368** (1992), 625–654.
- [100] S. Dalley, C. V. Johnson, and T. R. Morris, *Nonperturbative two-dimensional quantum gravity*, Nucl. Phys. **B368** (1992), 655–670.
- [101] M. Rothstein, *The Structure of supersymplectic supermanifolds*, in: Differential geometric methods in theoretical physics. Proceedings, 19th International Conference, Rapallo, Italy, June 19-24, 1990, pp. 331–343.
- [102] T. T. Voronov, *On volumes of classical supermanifolds*, Sbornik: Mathematics **207** (2016), no. 11, 1512.

- [103] E. Witten, *Notes on supermanifolds and integration*, arXiv:1209.2199 [hep-th].
- [104] E. Brézin and S. Hikami, *Random supermatrices with an external source*, JHEP **08** (2018) 086.
- [105] G. McShane, *Simple geodesics and a series constant over Teichmüller space*, Invent. Math. **132** (1998), 607–632.
- [106] L. Crane and J. M. Rabin, *Super Riemann surfaces: Uniformization and Teichmüller theory*, Commun. Math. Phys. **113** (1988) 601.
- [107] S. N. Dolgikh, A. A. Roslyi, and A. S. Schwarz, *Supermoduli spaces*, Commun. Math. Phys. **135** (1990), 91–100.
- [108] M. A. Baranov, I. V. Frolov, and A. S. Shvarts, *Geometry of Two-dimensional Superconformal Field Theories*, Theor. Math. Phys. **70** (1987), 64–72. [Teor. Mat. Fiz. **70** (1987) 92].
- [109] I. C. H. Ip, R. C. Penner, and A. M. Zeitlin, *On Ramond decorations*, arXiv:1709.06207 [math.GT].
- [110] T. Nagao and K. Slevin, *Nonuniversal correlations for random matrix ensembles*, Journal of mathematical physics **34** (1993). no. 5, 2075–2085.
- [111] J. M. Maldacena, G. W. Moore, N. Seiberg, and D. Shih, *Exact vs. semiclassical target space of the minimal string*, JHEP **10** (2004) 020.
- [112] P. Zograf, *On the large genus asymptotics of Weil-Petersson volumes*, arXiv:0812.0544 [math-ph].
- [113] M. Mirzakhani, *Growth of Weil-Petersson volumes and random hyperbolic surface of large genus*, Journal of Differential Geometry **94** (2013), no. 2, 267–300.
- [114] M. Mirzakhani and P. Zograf, *Towards large genus asymptotics of intersection numbers on moduli spaces of curves*, Geometric and Functional Analysis **25** (2015), no. 4, 1258–1289, arXiv:1112.1151 [math-ph].
- [115] G. W. Moore, N. Seiberg, and M. Staudacher, *From loops to states in 2-D quantum gravity*, Nucl. Phys. **B362** (1991), 665–709.
- [116] N. Seiberg and D. Shih, *Branes, rings and matrix models in minimal (super)string theory*, JHEP **02** (2004) 021.

- [117] J. Polchinski, *String Theory. Vol. 2: Superstring Theory and Beyond*, Cambridge Monographs on Mathematical Physics, Cambridge University Press, (2007).
- [118] I. R. Klebanov, J. M. Maldacena, and N. Seiberg, *Unitary and complex matrix models as 1-d type 0 strings*, *Commun. Math. Phys.* **252** (2004), 275–323.
- [119] T. Takayanagi and N. Toumbas, *A Matrix model dual of type 0B string theory in two-dimensions*, *JHEP* **07** (2003) 064.
- [120] M. R. Douglas, I. R. Klebanov, D. Kutasov, J. M. Maldacena, E. J. Martinec, and N. Seiberg, *A New hat for the $c = 1$ matrix model*, [arXiv:hep-th/0307195](https://arxiv.org/abs/hep-th/0307195) [hep-th].
- [121] S. Dalley, C. V. Johnson, T. R. Morris, and A. Watterstam, *Unitary matrix models and 2-D quantum gravity*, *Mod. Phys. Lett.* **A7** (1992), 2753–2762.
- [122] C. Crnkovic, M. R. Douglas, and G. W. Moore, *Loop equations and the topological phase of multi-cut matrix models*, *Int. J. Mod. Phys.* **A7** (1992), 7693–7711.

INSTITUTE FOR ADVANCED STUDY
EINSTEIN DRIVE, PRINCETON, NJ 08540, USA
STANFORD INSTITUTE FOR THEORETICAL PHYSICS
STANFORD, CA 94305, USA
E-mail address: salguod@stanford.edu

INSTITUTE FOR ADVANCED STUDY
EINSTEIN DRIVE, PRINCETON, NJ 08540, USA
E-mail address: witten@ias.edu