Syddansk Universitet

Topological recursion for chord diagrams, RNA complexes, and cells in moduli spaces

Andersen, Jørgen Goul; chekhov, Leonid; Penner, Robert; Reidys, Christian; Sukowski, Piotr

Published in:
Nuclear Physics, Section B

DOI:
10.1016/j.nuclphysb.2012.09.012

Publication date:
2013

Citation for pulished version (APA):
Andersen, J. G., chekhov, L., Penner, R., Reidys, C., & Sukowski, P. (2013). Topological recursion for chord diagrams, RNA complexes, and cells in moduli spaces. Nuclear Physics, Section B, 866(3), 414-443. https://doi.org/10.1016/j.nuclphysb.2012.09.012

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 23. Mar. 2019

arXiv:1205.0658v1 [hep-th] 3 May 2012

CALT-68-2869

TOPOLOGICAL RECURSION FOR CHORD DIAGRAMS, RNA COMPLEXES, AND CELLS IN MODULI SPACES

JØRGEN E. ANDERSEN[1], LEONID O. CHEKHOV[3], R. C. PENNER[1,2,⋆], CHRISTIAN M. REIDYS[4], PIOTR SUŁKOWSKI[2,5,6]

[1] Center for Quantum Geometry of Moduli Spaces, Aarhus University, DK-8000 Århus C, Denmark
[2] Division of Physics, Mathematics and Astronomy, California Institute of Technology, Pasadena, California, 91125 USA
[3] Department of Theoretical Physics, Steklov Mathematical Institute, Moscow, 119991 Russia
[4] Department of Mathematics and Computer Science, University of Southern Denmark, DK-5230 Odense M, Denmark
[5] Institute for Theoretical Physics, University of Amsterdam, Science Park 904, 1090 GL, Amsterdam, The Netherlands
[6] Faculty of Physics, University of Warsaw, ul. Hoża 69, 00-681 Warsaw, Poland
⋆ Corresponding author: rpenner@caltech.edu

Abstract. We introduce and study the Hermitian matrix model with potential $V_{s,t}(x) = x^2/2 - stx/(1-tx)$, which enumerates the number of linear chord diagrams of fixed genus with specified numbers of backbones generated by s and chords generated by t. For the one-cut solution, the partition function, correlators and free energies are convergent for small t and all s as a perturbation of the Gaussian potential, which arises for $st = 0$. This perturbation is computed using the formalism of the topological recursion. The corresponding enumeration of chord diagrams gives at once the number of RNA complexes of a given topology as well as the number of cells in Riemann's moduli spaces for bordered surfaces. The free energies are computed here in principle for all genera and explicitly for genera less than four.

Keywords: chord diagrams, matrix model, Riemann's moduli space, RNA, topological recursion

Introduction

Consider a collection of $b \geq 1$ pairwise disjoint, oriented and labeled intervals lying in the real line $\mathbb{R} \subset \mathbb{C}$, each component of which is called a *backbone*. A *chord diagram* C on these backbones is comprised of a collection of $n \geq 0$ semi-circles called *chords* lying in the upper half plane whose endpoints lie at distinct interior points of the backbones *so that the resulting diagram is connected*.

This description of C with its chords in the upper half plane determines a corresponding *fatgraph*, namely, a graph in the usual sense of the term together with cyclic orderings on each collection of half-edges incident on a common vertex. The fatgraph C, in turn, determines an associated *skinny surface* $\Sigma(C)$ with boundary by replacing each backbone with a rectangle to which are attached further semi-circular rectangles respecting orientation, one for each chord of C. In particular, $\Sigma(C)$ contains C as a deformation retract; see Figure 1 for an example, where the chosen under/overcrossings of the rectangles are immaterial to the fatgraph structure.

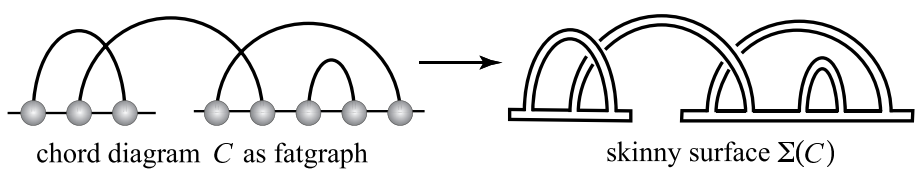

Figure 1. From chord diagram to skinny surface.

A chord diagram C therefore naturally determines an oriented surface $\Sigma(C)$, which is connected since C is assumed to be connected. Thus, $\Sigma(C)$ is characterized up to homeomorphism by its genus $g \geq 0$ and number $r \geq 1$ of boundary components, which we may equivalently associate to C itself, and the Euler characteristic of C or $\Sigma(C)$ is given by $b - n = 2 - 2g - r$. This last relation can be used in particular to find the number of boundary components of the surface $\Sigma(C)$ by inspection and then infer the genus. For example in Figure 1, we have $b = 2$ backbones, $n = 4$ chords and $r = 4$ boundary components, which implies that the corresponding surface has genus $g = 0$.

Let $c_{g,b}(n)$ denote the number of isomorphism classes of chord diagrams of genus g with n chords on b labeled backbones whose generating functions

$$C_{g,b}(z) = \sum_{n \geq 0} c_{g,b}(n)\ z^n, \text{ for } g \geq 0, \tag{1}$$

are of our central interest here. We shall recursively calculate them using the topological recursion [15, 16, 21] of a Hermitian one-matrix model

$$Z = \int DH\ e^{-N \mathrm{tr} V(H)} = \exp \sum_{g=0}^{\infty} N^{2-2g} F_g, \tag{2}$$

where N denotes size of matrices, for a particular potential

$$V(x) = \frac{x^2}{2} - \frac{stx}{1-tx} \tag{3}$$

generalizing the Gaussian for $st = 0$. The free energy in genus g, which we compute, is given by

$$F_g(s,t) = const + \sum_{b\geq 1} \frac{s^b}{b!}\, C_{g,b}(t^2), \text{ for } g \geq 0, \tag{4}$$

where the constant terms reproduce the Gaussian free energies given by $\frac{B_{2g}}{2g(2g-2)}$ according to [24] in each genus $g \geq 2$ with appropriate modifications for $g = 0$ or 1, and where B_{2g} denote Bernoulli numbers. The extra factor $b!$ arises because $C_{g,b}(n)$ counts chord diagrams with labeled backbones as opposed to unlabeled in the topological recursion, where a permutation of backbones must nevertheless preserve backbone orientations as we will see.

As has long been known, cf. [6, 35, 27], Hermitian matrix models are well suited to various computations of combinatorial fatgraph generating functions which are determined by corresponding potentials. In section 2, we derive the specific matrix model potential $V(x) = stx/(1-tx)$ which encodes the solution to our particular combinatorial problem. This is a two-parameter perturbation of the Gaussian, and the parameters s and t in the expansion of the partition function $Z_{s,t}$ are essentially generating parameters for the numbers of backbones and chords respectively. The entire solution of the model is encoded in the free energies F_g in (2), which arise in the large N expansion of the matrix integral. Provided the 't Hooft parameter T is non-zero (and we shall ultimately take it to be $T = 1$ to restrict to the model of the form (2)), we show there is a unique continuous one-cut extension of the Gaussian which converges for small t and all s. It is this solution that we compute here and show reproduces more elementary enumerative results.

Let us explain in more detail what the topological recursion is and how we can use it to determine the solution of the above matrix model. For a general Hermitian matrix model in its formal large N expansion, one can write down so-called loop equations, which are Ward identities or Schwinger-Dyson equations for certain multi-linear correlators $W_n^{(g)}(p_1, \ldots, p_n)$ generalizing the resolvent. The leading order equation among those identities specifies a so-called spectral curve, i.e., an algebraic curve on which the resolvent is well defined. It also turns out that all correlators $W_n^{(g)}(p_1, \ldots, p_n)$ and loop equations they satisfy can be encoded entirely in terms of this spectral curve. These loop equations can be solved in a recursive way [15, 16, 21], and in this manner, free energies F_g (for $g \geq 2$) are completely determined by correlators $W_1^{(g)}(p)$.

We stress that this entire procedure requires just the knowledge of the spectral curve and not the details of the matrix model from which this curve was derived. An important achievement of Eynard and Orantin [21]

was to realize that one can use the recursive solution of loop equations to assign correlators $W_n^{(g)}(p_1,\dots,p_n)$ and F_g to an arbitrary algebraic curve, not necessarily of matrix model origin. On the other hand, it is guaranteed that F_g computed for the spectral curve of a matrix model reproduce the free energies. For various interesting applications of this topological recursion to other problems related to moduli spaces of Riemann surfaces or algebraic curves, cf. [8, 9, 10, 11, 13, 24, 20, 21, 22, 31, 32, 33, 44].

It was already essentially known that in terms of the cut endpoints $a<b$, the eigenvalue density $\rho(x)$ and the spectral curve $y=M(x)\sqrt{(x-a)(x-b)}$, one can compute the leading free energies F_0 and F_1 of (2), and we follow this procedure for our particular matrix model to derive the multiplicities $c_{g,b}(n)$ of chord diagrams, for $g=0,1$, in appendix A. The genus zero free energy can be found from a solution to the variational problem [12, 29], which leads to the following form of the effective potential

$$F_0=-\frac{T}{2}\Big(\int dx\rho(x)V(x)-\lambda t_0\Big), \tag{5}$$

where λ is a Lagrange multiplier enforcing preservation of the number of eigenvalues $t_0=\int\rho(x)dx$. The genus one free energy can be computed as in [3] (for generalization to more cuts see [14]):

$$F_1=-\frac{1}{24}\log\left(M(a)M(b)(a-b)^4\right). \tag{6}$$

Explicit results for genus zero and one in the model with potential (3) are presented in appendix A. These results reproduce known answers (up to three backbones, as given in (10), (12) and (18)), and more importantly provide generating functions for arbitrary numbers of chords and backbones. For example, we determine that generating functions of diagrams in genus zero and one, on four backbones, with an arbitrary number of chords take the form

$$\begin{aligned} C_{0,4}(z)&=\frac{24z^3(3+18z+8z^2)}{(1-4z)^7}=72z^3+2448z^4+\dots\\ C_{1,4}(z)&=\frac{24z^5(715+7551z+12456z^2+2096z^3)}{(1-4z)^{10}}=17160z^5+\dots \end{aligned} \tag{7}$$

To illustrate the combinatorial complexity of chord diagrams and appreciate precisely the generating functions which we derive, we show in appendix C by explicit enumeration that $c_{0,4}(3)=72$ and $c_{0,4}(4)=2448$ as predicted by (7).

In this paper, apart from the side-note of determining the general F_0 (by saddle-point methods) and F_1 (due to Chekhov [14]) and analyzing them for our model in appendix A, our main result is a procedure based on topological recursion for calculating any desired one-cut free energy F_g, for any $g\geq 2$,

on a sufficiently large computer. We implement it to find, for example,

$$
\begin{aligned}
F_2 = &-\frac{t^4(1-\sigma)^2}{240\delta^4(1-\delta-4\sigma+3\sigma^2)^5(1+\delta-4\sigma+3\sigma^2)^5}\times \\
&\times\Big(160\delta^4(1-3\sigma)^4(1-\sigma)^6-80\delta^2(1-3\sigma)^6(1-\sigma)^8 \\
&+16(1-3\sigma)^8(1-\sigma)^{10}+\delta^{10}(-16+219\sigma-462\sigma^2+252\sigma^3) \\
&+10\delta^6(1-3\sigma)^2(1-\sigma)^4(-16-126\sigma-423\sigma^2+2286\sigma^3-2862\sigma^4+1134\sigma^5) \\
&+5\delta^8(1-\sigma)^2(16+189\sigma-2970\sigma^2+9549\sigma^3-11286\sigma^4+4536\sigma^5)\Big),
\end{aligned}
$$

where $\sigma = t(a+b)/2$ and $\delta = t(a-b)/2$. This can be solved exactly in one of s, t and perturbatively in the other to extract specific $C_{g,b}(z)$ as we will explain. These methods of computation apply in all higher genera as well, and our similar but longer expression for F_3 is given in (63). We again reproduce known results (up to three backbones, given in (12) and (18)) and then find generating functions for arbitrary numbers of chords and backbones. For instance as consequences of our computations, we find:

$$
\begin{aligned}
C_{2,4}(z) &= \frac{144z^7}{(1-4z)^{13}}(38675+620648z+2087808z^2 \\
&\quad +1569328z^3+134208z^4), \\
C_{3,4}(z) &= \frac{48z^9}{(1-4z)^{16}}(53416125+1194366915z+6557325096z^2 \\
&\quad +10738411392z^3+4580024832z^4+236239616z^5), \\
C_{2,5}(z) &= \frac{144z^8}{(1-4z)^{\frac{31}{2}}}(2543625+62424520z+375044396z^2 \\
&\quad +671666053z^3+314761848z^4+18335696z^5), \\
C_{3,5}(z) &= \frac{720z^{10}}{(1-4z)^{\frac{37}{2}}}(360380790+11275076865z+95744892585z^2 \\
&\quad +282797424880z^3+291167707410z^4+85497242928z^5 \\
&\quad +3218434848z^6).
\end{aligned}
$$

Let us also comment on the profound import and wide scope of these results. The numbers $c_{g,b}(n)$ are of significance in computational biology because they describe the possible non-trivial topological types of complexes of several interacting RNA molecules as follows. Each backbone is identified with the sugar-phosphate backbone (hence the terminology) of a single RNA molecule oriented from its 5' to 3' end. If two nucleic acids comprising the b RNA molecules participate in a Watson-Crick basepair, then we add a corresponding chord taking care that chord endpoints in each backbone

occur in the correct order corresponding to the primary structure, i.e., the word in the four-letter alphabet of nucleic acids that determines the RNA molecule. In this way, a complex of interacting RNA molecules determines a chord diagram, and we demand that it is connected in order to guarantee an appropriate non-triviality of the interaction. Note in particular that nucleotides not participating in basepairs play no role in this model, i.e., there are no isolated vertices.

The genus of the diagram determines the topological complexity of the interaction. This genus filtration has been profitably employed in a number of studies [34, 38, 42, 7, 23, 18] as well as in folding algorithms [41, 39] in the special case of a single RNA molecule (i.e., $b = 1$), which will be further discussed in section 1.2. Likewise, a folding algorithm for two interacting RNA molecules (i.e., $b = 2$) pertinent to antisense RNA for instance has been studied in [4].

At the same time, a simple transform of these numbers $c_{g,b}(n)$ count a sub-class of chord diagrams called "shapes". These are discussed in the next section and give the number of cells in the ideal cell decomposition [37] of Riemann's moduli space for a surface of genus $g \geq 0$ with $b \geq 1$ boundary components provided $2g + b > 2$.

The numbers computed here using the topological recursion are therefore at once of significance in computational biology and in geometry thus representing a remarkable confluence of biology, mathematics and physics.

This paper is organized as follows. There is background material in the next section including computations with the Gaussian potential itself for chord diagrams on at most three backbones. In section 2, we introduce our new model and determine the form of its potential (3). In order to solve this model, we first review some general relevant properties of matrix models in section 3 and introduce the formalism of topological recursion in section 4. Finally in section 5, we solve the matrix model (2) with the potential (3): we begin by determining its spectral curve and then compute the correlators $W_n^{(g)}(p_1, \ldots, p_n)$ and free energies F_g assigned to it through the topological recursion formalism. As the potential (3) depends on parameters s (which generates numbers of backbones) and t (which generates numbers of chords), also the spectral curve, and then correlators and free energies depend on these two parameters as well. In appendix A, we present the computations of F_0 and F_1, which are performed independently of the topological recursion, and in appendix B, we write down the (perturbative) solution of the equations determining cut endpoints in our model. In the final appendix C, we directly enumerate particular classes of chord diagrams just to confirm that indeed $c_{0,4}(3) = 72$ and $c_{0,4}(4) = 2448$. This enumeration also gives a sense of the simplest cases of objects we are counting here and can be read first if desired.

1. Background

1.1. **Chord diagrams, seeds and shapes.** Two chords γ and γ' in a chord diagram C with respective endpoints x, y and x', y' are said to be *parallel* if x, x', as well as y, y', lie in a common backbone with no chord endpoints in between, where $x < x'$ and $y' < y$. Parallelism generates an equivalence relation whose equivalence classes are called *stacks*.

Suppose the endpoints x, y of a chord γ lie in a common backbone β. If there are no chord endpoints between x, y, then γ is called a *pimple* on β, and if all chord endpoints in β lie between x and y, then γ is called a *rainbow* on β.

A *seed* is a chord diagram where every stack has cardinality one so that each pimple is a rainbow, and a seed is a *shape* provided every backbone has a rainbow. In particular for a chord diagram of genus zero, an innermost chord with both endpoints on a single backbone is necessarily a pimple, so for example, the only seeds of genus zero on one backbone are the empty diagram with no chords and the diagram with a single chord given by the rainbow, and only the latter is also a shape.

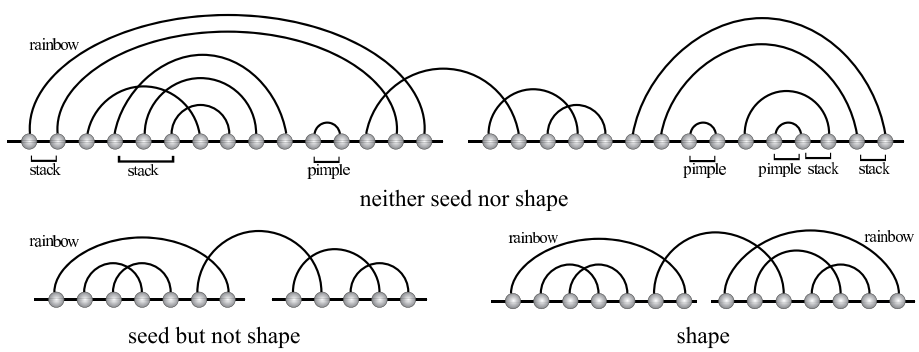

Figure 2. Stacks, pimples, rainbows, shapes, seeds.

Proposition 1.1. *Other than the case $g = 0, b = 1$ (which was just discussed), a seed of genus g on b backbones with n chords must satisfy the inequalities $2g + b - 1 \leq n \leq 6g - 6 + 5b$, and these constraints are sharp.*

Proof. The lower bound on n follows from the fact that $2g+b-1 \leq n+1-r$ according to Euler characteristic considerations together with the obvious constraint that $r \geq 1$. Conversely, a seed which saturates this lower bound necessarily has $r = 1$.

If the skinny surface associated to a seed has more than one boundary component, then there must be a chord with different boundary components on its two sides; removing this chord decreases r by exactly one while preserving g again from Euler characteristic considerations. Define the "length" of a boundary component to be the number of chords it traverses counted with multiplicity. If there are ν_ℓ boundary components of length ℓ, then $2n =$

$\sum_\ell \ell\nu_\ell$ since each side of each chord is traversed exactly once in the boundary. It follows that $2n = 2(b+r+2g-2) \geq \nu_1+2\nu_2+3(r-\nu_1-\nu_2) \geq 3r-2b$ since $2\nu_1+\nu_2 \leq 2b$ (except in the excluded case for which $2\nu_1+\nu_2 \leq 4b$), so $4(b+g-1) \geq r$. There can thus be at most $4g+4b-5$ such removals of chords to produce a seed with $r=1$ proving the upper bound on n. □

Let $c_{g,b}(n)$, $s_{g,b}(n)$ and $t_{g,b}(n)$, respectively, denote the number of isomorphism classes of (connected) chord diagrams, seeds and shapes of genus $g \geq 0$ with $n \geq 0$ chords on $b \geq 1$ backbones. In each alphabetic case of $X = C, S, T$, we define corresponding generating functions

$$X_{g,b}(z) = \sum_{n\geq 0} x_{g,b}(n)\ z^n, \tag{8}$$

and let

$$X_g(z) = X_{g,1}(z). \tag{9}$$

Note that whereas $C_{g,b}(z)$ is a formal power series, $S_{g,b}(z)$ and $T_{g,b}(z)$ are polynomials by Proposition 1.1.

In particular, $C_0(z)$ is the generating function for the Catalan numbers, i.e., $c_0(n) = c_{0,1}(n)$ is the number of triangulations of a fixed polygon with $n+2$ sides. They evidently satisfy $c_0(n+1) = \sum_{i=0}^n c_0(i)c_0(n-i)$ with basis $c_0(n) = 1$, which gives $C_0(z) = 1 + z[C_0(z)]^2$, whence

$$C_0(z) = \frac{1-\sqrt{1-4z}}{2z}. \tag{10}$$

Theorem 1.2. *The generating functions for seeds and chord diagrams are related by*

$$C_{g,b}(z) = [C_0(z)]^b\ S_{g,b}\left(\frac{C_0(z)-1}{2-C_0(z)}\right),$$

$$S_{g,b}(z) = \left[\frac{z+1}{1+2z}\right]^b C_{g,b}\left(\frac{z(1+z)}{(1+2z)^2}\right).$$

Furthermore, the generating functions for seeds and shapes are related by

$$(1+z)^b\ T_{g,b}(z) = z^b\ S_{g,b}(z).$$

Proof. Writing simply C_0 for $C_0(z)$ and using $C_0 - 1 = zC_0^2$, we find

$$\frac{C_0-1}{2-C_0} = \frac{C_0-1}{1-(C_0-1)} = zC_0^2\sum_{i\geq 0}(zC_0^2)^i = \sum_{j\geq 1}(zC_0^2)^j.$$

As an argument of $S_{g,b}$, the jth term in the sum corresponds to inflating a single arc in a seed to a stack of cardinality $j \geq 1$ as well as inserting a genus zero diagram immediately preceding each of the resulting $2j$ chord endpoints. There is yet another factor C_0 arising from the insertion of a genus zero diagram following the last endpoint of the seed on each backbone

accounting for the further factor C_0^b. The resulting chord diagram is connected if and only if the seed itself is connected, and this proves the first formula.

For the second formula, direct calculation shows that $z = \frac{u(1+u)}{(1+2u)^2}$ inverts the expression $u = \frac{C_0(z)-1}{2-C_0(z)} = \frac{1-\sqrt{1-4z}}{2\sqrt{1-4z}}$, so the first formula reads

$$S_{g,b}(u) = \left[C_0\left(\frac{u(1+u)}{(1+2u)^2}\right)\right]^{-b} C_{g,b}\left(\frac{u(1+u)}{(1+2u)^2}\right).$$

Further direct computation substituting $z = \frac{u(1+u)}{(1+2u)^2}$ into $C_0(z) = \frac{1-\sqrt{1-4z}}{2z}$ shows that $C_0\big(\frac{u(1+u)}{(1+2u)^2}\big) = \frac{1+2u}{1+u}$, and the expression for $S_{g,b}$ follows.

The third formula is truly elementary since a shape is by definition simply a seed together with a rainbow on each backbone. □

Seeds are useful insofar as chord diagrams as well as their generalizations allowing isolated vertices which arise in practice can be enumerated from them, cf. [41, 40] for instance. Shapes are important because of the following fundamental observation from [5, 37]:

Theorem 1.3. *Provided $2g + b > 2$, the set of shapes of genus $g \geq 0$ on $b \geq 1$ backbones with its natural partial order under inclusion is isomorphic to the ideal cell decomposition of Riemann's moduli space of a surface of genus g with b boundary components.*

Proof. Collapsing each backbone to a distinct vertex produces from a shape a fatgraph where each vertex has a loop connecting consecutive half-edges arising from the rainbow. The dual fatgraph of this fatgraph arises by interchanging the roles of vertices and boundary components. The collection of duals of shapes of genus g on b backbones gives precisely those fatgraphs that arise in the Penner-Strebel ideal cell decomposition of the associated Riemann moduli space, where the collapsed rainbows are dual to the tails discussed in [37]. Removal of chords corresponds to contraction of dual fatgraph edges. □

In light of Theorem 1.2, the explicit calculation of the generating function for chord diagrams we perform here thus gives as a consequence also the numbers of cells of given dimension in Riemann's moduli spaces of bordered surfaces.

1.2. **The special case of one backbone.** For one backbone, there is the remarkable formula of Harer-Zagier [26], which we discuss here, that arose in the calculation of the virtual Euler characteristics of Riemann moduli spaces for punctured surfaces and in principle solves for the various $c_g(n) = c_{g,1}(n)$. In effect, the Penner matrix model [35] directly computed the former without recourse to the latter. There is a large literature on the Harer-Zagier formula, cf. [28] and the references therein including several derivations of it. Here is this beautiful and striking formula:

Theorem 1.4. *We have the identity*

$$1+2\sum_{n\geq 0}\sum_{2g\leq n}\frac{c_g(n)N^{n+1-2g}}{(2n-1)!!}\ z^{n+1}=\left(\frac{1+z}{1-z}\right)^N.$$

Furthermore, the $c_g(n)$ satisfy the recursion

$$(n+1)\,c_g(n)=2(2n-1)\,c_g(n-1)+(2n-1)(n-1)(2n-3)\,c_{g-1}(n-2).$$

This recursion directly translates into the ODE

$$z(1-4z)\frac{d}{dz}C_g(z)+(1-2z)C_g(z) \quad = \quad \Phi_{g-1}(z), \tag{11}$$

where

$$z^{-2}\Phi_{g-1}(z)=4z^3\frac{d^3}{dz^3}C_{g-1}(z)+24z^2\frac{d^2}{dz^2}C_{g-1}(z)+27z\frac{d}{dz}C_{g-1}(z)+3C_{g-1}(z)$$

with initial condition $C_g(0)=0$. From the ODE (11) and direct calculation follow several results:

Corollary 1.5. *For any $g\geq 1$, the generating function $C_g(z)$ is expressed as*

$$C_g(z)=\frac{P_g(z)}{(1-4z)^{3g-\frac{1}{2}}}, \tag{12}$$

where $P_g(z)$ is an integral polynomial divisible by z^{2g} but no higher power and of degree at most $(3g-1)$ with $P_g(1/4)\neq 0$.

The first few $P_g(z)$ are given as follows:

$$\begin{aligned}
P_1(z) &= z^2,\\
P_2(z) &= 21z^4\ (z+1)\\
P_3(z) &= 11z^6\ \left(158\,z^2+558\,z+135\right),\\
P_4(z) &= 143z^8\left(2339\,z^3+18378\,z^2+13689\,z+1575\right),\\
P_5(z) &= 88179z^{10}\ \left(1354\,z^4+18908\,z^3+28764\,z^2+9660\,z+675\right).
\end{aligned}$$

It is natural to speculate that the polynomials $P_g(z)$ themselves solve an enumerative problem and that there may be a purely combinatorial topological proof of Theorem 1.4 based on a construction that in some way "inflates" these structures using genus zero diagrams, to wit

$$\begin{aligned}
C_g(z) &= P_g(z)(\sqrt{1-4z})^{1-6g}\\
&= P_g(z)\left(\frac{C_0(z)}{2-C_0(z)}\right)^{6g-1}\\
&= P_g(z)(C_0(z))^{6g-1}\left(1+(C_0(z)-1)+(C_0(z)-1)^2+\cdots\right)^{6g-1}\\
&= P_g(z)(C_0(z))^{6g-1}\left(1+z(C_0(z))^2+z^2(C_0(z))^4+\cdots\right)^{6g-1}.
\end{aligned}$$

Another calculational consequence of the ODE (11) is an explicit recursion for the numbers of seeds:

Corollary 1.6. *We have $S_1(z) = z^2 + 3\,z^3 + 3\,z^4 + z^5$, and for any $g > 1$, the following five-term recursion for the number of seeds:*

$$(n+1)\,s_g(n) \;=\; \sum_{h=2}^{6} g_h(n)\,s_{g-1}(n-h), \quad \textit{where} \tag{13}$$

$$\begin{aligned}
g_2(n) &= (2n-1)(n-1)(2n-3)\\
g_3(n) &= 2(4n-11)(n-1)(2n-3)\\
g_4(n) &= (24n^3 - 180n^2 + 411n - 285)\\
g_5(n) &= (16n^3 - 156n^2 + 464n - 414)\\
g_6(n) &= 4(n-2)(n-4)(n-6).
\end{aligned}$$

The first few seed polynomials are thus given by:

$$\begin{aligned}
S_2(z) &= 21\,z^4 + 210\,z^5 + 840\,z^6 + 1785\,z^7 + 2205\,z^8 + 1596\,z^9\\
&\quad +630\,z^{10} + 105\,z^{11}\\
S_3(z) &= 1485\,z^6 + 28413\,z^7 + 225379\,z^8 + 1006929\,z^9 + 2862783\,z^{10}\\
&\quad +5496645\,z^{11} + 7325703\,z^{12} + 6812289\,z^{13} + 4348344\,z^{14}\\
&\quad +1819818\,z^{15} + 450450\,z^{16} + 50050\,z^{17}\\
S_4(z) &= 225225\,z^8 + 6687252\,z^9 + 83809440\,z^{10} + 605300410\,z^{11}\\
&\quad +2867032740\,z^{12} + 9542753220\,z^{13} + 23243924704\,z^{14}\\
&\quad +42438380985\,z^{15} + 58817592405\,z^{16} + 62093957640\,z^{17}\\
&\quad +49660516620\,z^{18} + 29612963952\,z^{19} + 12768025270\,z^{20}\\
&\quad +3763479720\,z^{21} + 678978300\,z^{22} + 56581525\,z^{23}\\
S_5(z) &= 59520825\,z^{10} + 2458871415\,z^{11} + 43395443091\,z^{12} + 447114000333\,z^{13}\\
&\quad +3067654998408\,z^{14} + 15065930976096\,z^{15} + 55383949540920\,z^{16}\\
&\quad +156912145081692\,z^{17} + 349331909457531\,z^{18} + 618798832452801\,z^{19}\\
&\quad +878354703383157\,z^{20} + 1001499812704755\,z^{21} + 915317216039226\,z^{22}\\
&\quad +665807616672198\,z^{23} + 380384810323518\,z^{24} + 166997907962886\,z^{25}\\
&\quad +54384464384250\,z^{26} + 12376076963250\,z^{27} + 1756856351250\,z^{28}\\
&\quad +117123756750\,z^{29}
\end{aligned}$$

Remark 1.7. [25] With further work starting from this recursion, one can show that the number of seeds is log concave, i.e., for $2g+1 \le n \le 6g-2$, we have

$$s_g(n)^2 \ge s_g(n+1)\,s_g(n-1).$$

In particular, it follows that the sequence $s_g(2g), s_g(2g+1), \ldots, s_g(6g-1)$ is unimodal.

1.3. **The Gaussian resolvents for at most 3 backbones.** Provided the number b of backbones is less than four, the loop equations [2, 19] for the Gaussian potential permit the computation of the $c_{g,b}(n)$ in terms of $c_g(n)$ as we discuss here. This in part explains an important aspect of the computations of the previous section. See also [30].

For any $s \geq 1$, the *s-resolvent* or *s-point function* of the potential $V(x)$ is defined to be

$$\omega_s(x_1,\ldots,x_s) = N^{s-2}\ \langle \mathrm{tr}\ \frac{1}{x_1 - H}\ \cdots\ \mathrm{tr}\ \frac{1}{x_s - H}\rangle_{\mathrm{conn}}, \tag{14}$$

where $\langle \Phi \rangle_{conn}$ denotes the part of the correlator $\frac{1}{Z}\int DH\ e^{-N\mathrm{tr}V(H)}\Phi(H)$ arising via Wick's Theorem from connected diagrams computed as a formal power series using $\mathrm{tr}\ \frac{1}{x-H} = \sum_{k\geq 0} \mathrm{tr}\ \frac{H^k}{x^{k+1}}$, where the integral is over the $N \times N$ Hermitian matrices with their Haar measure

$$DH = \left(\Lambda_{1\leq i<j\leq N}\ dReH_{ij} \wedge dImH_{ij}\ \right) \wedge \left(\Lambda_{i=1}^{n} dH_{ii}\right) \tag{15}$$

and with the partition function $Z = \int DH\ e^{-N\mathrm{tr}V(H)}$.

For a polynomial potential $V(x) = \sum_{k=1}^{d} t_k x^k$ for convenience (our methods apply more generally whenever the potential has rational derivative), we find that

$$\begin{aligned}\omega_s(x_1,\ldots,x_s) &= \sum_{g\geq 0} N^{-2g} \sum_{n_1,\ldots,n_s\geq 0} \frac{1}{x_1^{n_1+1}\cdots x_s^{n_s+1}} \\ &\qquad \sum_{\ell_1,\ell_3,\ldots,\ell_d\geq 0} \prod_{k=1,k\neq 2}^{d} (-t_k)^{\ell_k} \omega^{(g)}_{n_1,\ldots,n_s}(\{\ell_1,\ell_3,\ldots,\ell_d\}) \\ &= \sum_{g\geq 0} N^{-2g}\ \omega_s^{(g)}(x_1,\ldots,x_s),\end{aligned}$$

where $\omega^{(g)}_{n_1,\ldots,n_s}(\{\ell_1,\ell_3,\ldots,\ell_d\})$ denotes the number of isomorphism classes of connected chord diagrams of genus g on s labeled oriented backbones, with respective numbers $n_1,\ldots,n_s$ of incident half-chords, plus ℓ_k unlabeled oriented backbones with k incident half-chords, for $k \geq 1$.

Now specializing to the Gaussian potential $V(x) = \frac{1}{2}x^2$ and letting $\omega_s(x) = \omega_s(x,\ldots,x)$ for convenience, we find

$$\omega_s(x) = \sum_{g\geq 0} N^{-2g}\ \omega_s^{(g)}(x) = x^{-s}\ \sum_{g\geq 0} N^{-2g}\ C_{g,s}(x^{-2})$$

in the earlier notation since $C_{g,b}(z^{-2}) = z^s \omega_b^{(g)}(z)$. The first Schwinger-Dyson equation or Ward identity with equal arguments in this case reads $\omega_1^2(x) - x\omega_1(x) + 1 + N^{-2}\ \omega_2(x) = 0$, from which we solve for the 2-point function with equal arguments in terms of the 1-point function:

$$\omega_2(x) = N^2\{x\omega_1(x) - \omega_1^2(x) - 1\}. \tag{16}$$

Likewise, $x\omega_2(x) = 2\omega_1(x)\omega_2(x) + \frac{1}{2}\omega_1''(x) + N^{-2}\omega_3(x)$ gives the 3-point function with equal arguments

$$\omega_3(x) = N^2\{x\omega_2(x) - \frac{1}{2}\omega_1''(x) - 2\omega_1(x)\omega_2(x)\} \tag{17}$$

in terms of $\omega_1(x)$ and its derivatives.

Remark 1.8. One cannot continue this recursion to express higher resolvents $\omega_s(x)$, for $s \geq 4$, with equal arguments in terms of $\omega_1(x)$ alone because higher Schwinger-Dyson equations include products of traces (for example, $\omega_4(x)$ involves $\langle \mathrm{tr}\ \frac{1}{(x-H)^3}\ \mathrm{tr}\ \frac{1}{x-H}\rangle_{conn}$) which cannot be evaluated in this manner. We wonder if this is perhaps related to the behavior [36] of posets of chord diagrams: in the case of genus zero for s backbones with $s = 1, 2, 3$, the geometric realization of this poset is homeomorphic to a sphere; for $s = 4$, it is homeomorphic to a simply connected manifold which is not a sphere; and for $s > 4$, it is an increasingly singular non-manifold.

Theorem 1.9. *For any $g \geq 0$, we have*

$$C_{g,2}(x) = \frac{P_{g,2}(x)}{(1-4x)^{3g+2}} \text{ and } C_{g,3}(x) = \frac{P_{g,3}(x)}{(1-4x)^{3g+5-\frac{1}{2}}}, \tag{18}$$

where

$$\begin{aligned}
P_{g,2}(x) &= x^{-1}P_{g+1}(x) - \sum_{h=1}^{g} P_h(x)P_{g+1-h}(x) \text{ and} \\
P_{g,3}(x) &= x^{-1}P_{g+1,2}(x) - 2\sum_{h=1}^{g+1} P_h(x)P_{g+1-h,2}(x) \\
&\quad - P_{g+1}(x)\Big[1 + (42+60g)x + (96+336g+288g^2)x^2\Big] \\
&\quad - P_{g+1}'(x)(1-4x)\Big[5x + (20+48g)x^2\Big] \\
&\quad - P_{g+1}''(x)(1-4x)^2\Big[2x^2\Big].
\end{aligned}$$

Proof. It follows from (16) that $\omega_2(z)$ is given by

$$\sum_{g\geq 0} N^{-2g}\omega_2^{(g)}(z) = N^2\Big\{z\sum_{g\geq 0} N^{-2g}\omega_1^{(g)}(z) - \big[\sum_{g\geq 0} N^{-2g}\omega_1^{(g)}(z)\big]^2 - 1\Big\}.$$

Extracting the coefficient of N^{-2g}, we conclude that

$$\omega_2^{(g)}(z) = z\omega_1^{(g+1)}(z) - \sum_{h=0}^{g+1} \omega_1^{(h)}(z)\omega_1^{(g+1-h)}(z),$$

i.e.,

$$z^{-2}C_{g,2}(z^{-2}) = zz^{-1}C_{g+1}(z^{-2}) - \sum_{h=0}^{g+1} z^{-2}C_h(z^{-2})C_{g+1-h}(z^{-2}).$$

Setting $x = z^{-2}$, we find

$$\begin{aligned} C_{g,2}(x) &= x^{-1}C_{g+1}(x) - \sum_{h=0}^{g+1} C_h(x)C_{g+1-h}(x) \\ &= x^{-1}C_{g+1}(x) - \sum_{h=1}^{g} C_h(x)C_{g+1-h}(x) - 2C_0(x)C_{g+1}(x), \end{aligned}$$

and rewriting this using (12) gives

$$\begin{aligned} C_{g,2}(x) &= \left\{\frac{1}{x} - \frac{1-\sqrt{1-4x}}{x}\right\}\frac{P_{g+1}(x)}{(1-4x)^{3g+3-\frac{1}{2}}} \\ &\quad - \frac{1}{(1-4x)^{3g+2}}\sum_{h=1}^{g} P_h(x)P_{g+1-h}(x) \\ &= \frac{P_{g,2}(x)}{(1-4x)^{3g+2}} \end{aligned}$$

as was claimed.

Starting from (17), the analogous but slightly more involved computation produces the asserted expression for $C_{g,3}(x)$. □

For later comparison to our one-cut solution, we here record the first several such polynomials on two or three backbones in low genus:

$$\begin{aligned} P_{0,2}(x) &= x, \\ P_{1,2}(x) &= x^3(20x+21), \\ P_{2,2}(x) &= x^5\left(1696x^2+6096x+1485\right), \\ P_{3,2}(x) &= x^7\left(330560x^3+2614896x^2+1954116x+225225\right), \end{aligned}$$

$$\begin{aligned} P_{0,3}(x) &= 2x^2(3+4x), \\ P_{1,3}(x) &= 12x^4(45+207x+68x^2), \\ P_{2,3}(x) &= 6x^6(15015+137934x+197646x^2+27592x^3), \\ P_{3,3}(x) &= 8x^8(3132675+46335375x+143262162x^2 \\ &\quad +98362965x^3+7468348x^4), \end{aligned}$$

2. The potential $V(s,t,x)$

If C is a disjoint union of chord diagrams, then let

$b(C)$ denote its number of backbones,
$n(C)$ denote its number of chords,
$r(C)$ denote its number of boundary cycles,
$Aut(C)$ denote its automorphism group permuting oriented backbones, and
$g(C)$ denote the sum of genera of its components.

Furthermore, set $\mu_N = 2^{N/2}\pi^{N^2/2}$ and for any tuple $v_1, v_2, \ldots, v_K$ of non-negative integers, define

$$P_{v_1,\ldots,v_K}(s,t,N) = \frac{1}{\mu_N}\,\frac{1}{\prod_k v_k!}\int e^{-\mathrm{tr}\ H^2/2}\ \prod_k \Big(s\ \mathrm{tr}\ (tH)^k\Big)^{v_k}\ DH$$

for the integral as before over the $N \times N$ Hermitian matrices with respect to the Haar measure DH given in (15).

Lemma 2.1. *For any tuple $v_1, \ldots, v_K$, parameters s, t and natural number N, we have*

$$P_{v_1,\ldots,v_K}(s,t,N) = \sum \frac{N^{r(C)}\ s^{b(C)}\ t^{2n(C)}}{\#Aut(C)},$$

where the sum is over all isomorphism classes of disjoint unions of chord diagrams with v_k backbones having k incident half-chords, for $k = 1, \ldots, K$.

Proof. The special case $s = t = 1$ is entirely analogous to that of Theorem 2.1 of [35], where the replacement of $\frac{1}{k}$ tr H^k there by tr H^k here corresponds to replacing fatgraphs there by chord diagrams here, i.e., distinguishing one sector at each vertex of the fatgraph to represent the location of the backbone in the chord diagram kills the cyclic permutations about the vertices. The general case then follows easily since $b(C) = \sum_k v_k$ and $2n(C) = \sum_k kv_k$. □

A standard computation as on p. 49 of [35] thus gives

Theorem 2.2. *For $Z(s,t,N) = \frac{1}{\mu_N}\int e^{-N\ \mathrm{tr}\ V(s,t,H)}\ DH$ with potential*

$$V(s,t,H) = \frac{1}{2}H^2 - \frac{stH}{1-tH} = \frac{1}{2}H^2 - s\sum_{k\geq 1}(tH)^k, \tag{19}$$

we have

$$log\ Z(s,t,N) = \sum \frac{N^{2-2g(C)}\ s^{b(C)}\ t^{2n(C)}}{\#Aut(C)},$$

where the sum is over all (connected) chord diagrams C.

An amusing point, which however does not help with computations, is that the derivative of the logarithmic potential from [35]

$$\frac{d}{dx}\Big(x + log(1-x)\Big) = 1 + \frac{1}{x-1} = \frac{x}{x-1}$$

agrees with our potential here up to the Gaussian factor for $s = t = 1$. Let us also emphasize that the partition function in this section enumerates chord diagrams with unlabeled backbones in contrast to the considerations in section 1.2. (Vertices from insertions are labeled whereas vertices from the potential are not in Wick's Theorem.)

3. Matrix models and spectral curves: generalities

As we sketched in the introduction, the main idea of this paper is to solve the problem of counting chord diagrams (or RNA complexes or cells in Riemann's moduli space) using matrix model techniques and the topological recursion. We present this general formalism in this and the next section. We will apply it in section 5 to find the free energy expansion of the one-cut Hermitian matrix model with the potential given in (19).

Let us consider a general Hermitian matrix model

$$Z = \int DH\ e^{-\frac{1}{\hbar}\mathrm{tr}V(H)} = \ \exp\sum_{g=0}^{\infty}\hbar^{2g-2}F_g, \tag{20}$$

with the measure DH introduced in (15). This model is slightly more general than (2) as we replaced the explicit size of matrices N in the exponents by $1/\hbar$. In this section, we present how to find a spectral curve of such a model in the large N limit, with the so-called 't Hooft coupling

$$T = \hbar N = const \tag{21}$$

kept fixed. In the next section, we will explain how, from the knowledge of the spectral curve, one can use the topological recursion to determine free energies F_g. For a matrix model such as (20), the spectral curve and in consequence the F_g depend on T. To solve the combinatorial enumeration problem in (2), we can simply set $T = 1$.

The spectral curve arises from the leading order loop equation of a matrix model, and it can be characterized as a curve on which the resolvent is well-defined. The (all-order) resolvent is defined as a correlator

$$\omega_1(x) = \hbar\langle \mathrm{Tr}\frac{1}{x-M}\rangle = \sum_{g=0}^{\infty}\hbar^{2g}\omega_1^{(g)}(x). \tag{22}$$

The leading order term in this expansion $\omega_1^{(0)}(x)$ is often referred to simply as the *resolvent*. It is simply related to the density of eigenvalues $\rho(x)$ which becomes continuous in large N limit and has the following asymptotics for

large x

$$\omega_1^{(0)}(x) = \frac{1}{t_0}\int \frac{\rho(x')}{x-x'}dx' \underset{x\to\infty}{\sim} \frac{1}{x}, \tag{23}$$

where

$$t_0 = \int \rho(x)dx. \tag{24}$$

In this large N limit, the eigenvalues are distributed along compact intervals called cuts, along which in particular the above integrals are performed. Moreover, an inverse relation determines the density of eigenvalues as a discontinuity

$$\rho(x) = \frac{1}{2\pi i}\Big(\omega_1^{(0)}(x-i\epsilon) - \omega_1^{(0)}(x+i\epsilon)\Big) \tag{25}$$

in the resolvent along these cuts. The consistency of the above relations requires that

$$\omega_1^{(0)}(x-i\epsilon) + \omega_1^{(0)}(x+i\epsilon) = \frac{V'(x)}{T}. \tag{26}$$

The conditions (23) and (26) determine the resolvent $\omega_1^{(0)}(x)$ completely, and it is often possible to guess its form by requiring that these conditions are met. Equivalently, one can find the resolvent using the so-called Migdal formula. In case there is a single cut with endpoints denoted a and b, this formula reads

$$\omega_1^{(0)}(x) = \frac{1}{2T}\oint_{\mathcal{C}} \frac{dz}{2\pi i}\frac{V'(z)}{x-z}\frac{\sqrt{(x-a)(x-b)}}{\sqrt{(z-a)(z-b)}}, \tag{27}$$

where the contour $\mathcal{C}$ encircles the cut. Moreover, expanding the resulting expression around infinity and imposing the condition (23) provides a relation between 't Hooft coupling T and the cut endpoints. This one-cut resolvent will be sufficient for our computations in this paper, however, it is not hard to generalize the above formula to the case with many cuts having endpoints a_i and b_i; in this case, one only needs to replace the radicals in numerator and denominator of the integrand respectively by $\prod_i(x-a_i)(x-b_i)$ and $\prod_i(z-a_i)(z-b_i)$.

It is also convenient to encode the information about the resolvent $\omega^{(0)}(x)$ in a new variable

$$y = \frac{1}{2}V'(x) - T\omega(x) = \frac{T}{2}(\omega_1^{(0)}(x-i\epsilon) - \omega_1^{(0)}(x+i\epsilon)), \tag{28}$$

which turns out to be related to x by a polynomial equation

$$A(x,y) = 0. \tag{29}$$

This equation therefore defines an algebraic curve, which is called a spectral curve of a matrix model. As we will review below in the context of the topological recursion, one in fact needs to provide a parametric form of this equation, i.e., to represent it in terms of two functions $x = x(p)$ and

$y = y(p)$ of a parameter p living on the Riemann surface such that the equation $A(x(p), y(p)) = 0$ is satisfied identically. For the one-cut solution, the functions $x(p)$ and $y(p)$ are rational, while equation (28) takes the form

$$y = M(x)\sqrt{(x-a)(x-b)}, \tag{30}$$

where $M(x)$ is a rational function.

Example 3.1. Let us consider the Gaussian model with the potential $V(x) = \frac{1}{2}x^2$. The Migdal formula (27) results in

$$\omega_1^{(0)}(x) = \frac{1}{2T}\Big(x - \sqrt{(x-a)(x-b)}\Big) = \tag{31}$$

$$= \frac{a+b}{4T} + \frac{(a-b)^2}{16Tx} + \mathcal{O}(x^{-2}), \tag{32}$$

and from the first line, it is clear that (26) is satisfied. The expansion in the second line together with the condition (23) then imply that

$$a = -2\sqrt{T}, \qquad b = 2\sqrt{T}. \tag{33}$$

Also, introducing y as in (28), we find the following equation for the spectral curve

$$y^2 = \frac{1}{4}(x-a)(x-b). \tag{34}$$

Comparing with (30), we see that $M(x) = \frac{1}{2}$ in this case, and a parametric form of this algebraic curve can be given for example as follows

$$\begin{cases} x(p) = \frac{a+b}{2} + \frac{a-b}{4}\Big(p + \frac{1}{p}\Big) = -\sqrt{T}\Big(p + \frac{1}{p}\Big), \\ y(p) = \frac{a-b}{8}\Big(\frac{1}{p} - p\Big) = \frac{\sqrt{T}}{2}\Big(p - \frac{1}{p}\Big). \end{cases} \tag{35}$$

4. Topological recursion

The topological recursion introduced in [15, 16, 21] can be used to assign to an algebraic curve $A(x, y) = 0$ a series of multi-linear differentials $W_n^{(g)}(p_1, \ldots, p_n)$ and functions F_g called free energies. Free energies F_g are symplectic invariants, i.e., they are invariant with respect to transformations which preserve the two-form $dx \wedge dy$ up to a sign. The multi-differentials transform in an appropriate way under symplectic transformations, and they depend on a particular parametrization $x = x(p), y = y(p)$ of the algebraic curve. The differentials $W_n^{(g)}(p_1, \ldots, p_n)$ are defined recursively by equations which, by definition, coincide with loop equations of a Hermitian matrix model [2]; as shown in [19], all details of these equations depend only on the form of its underlying algebraic curve $A(x, y) = 0$. Thus, applying the topological recursion to the curve which is a spectral curve of some particular Hermitian matrix model such as (20), reproduces free energies F_g

and correlators $W_n^{(g)}(p_1,\dots,p_n)$. From the matrix model viewpoint, these correlators are defined as

$$\Big\langle \mathrm{tr}\Big(\frac{1}{x(p_1)-M}\Big)\cdots \mathrm{tr}\Big(\frac{1}{x(p_k)-M}\Big)\Big\rangle_{\mathrm{conn}} = \sum_{g=0}^{\infty}\hbar^{2g-2+n}\frac{W_n^{(g)}(p_1,\dots,p_n)}{dx(p_1)\dots dx(p_n)},$$

which generalizes the resolvent introduced in (22), and in particular, $W_1^{(g)}(x) = \omega_1^{(g)}(x)dx$. For $\hbar = N^{-1}$, the functional part of $W_n^{(g)}$ also agrees with $\omega_n^{(g)}$ introduced in (1.3).

The recursion relations for multilinear differentials $W_n^{(g)}$ are defined in terms of the following quantities. Firstly, these relations are expressed in terms of residues around the branch points

$$dx(p_i^*) = 0\,. \tag{36}$$

For a point p in the neighborhood of a branch point p_i^* there is a unique conjugate point $\bar{p}$, defined as

$$x(p) \;=\; x(\bar{p})\,. \tag{37}$$

Two other important ingredients are the differential 1-form

$$\omega(p) = \big(y(\bar{p}) - y(p)\big)dx(p) \tag{38}$$

and the Bergman kernel $B(p,q)$. The Bergman kernel $B(p,q)$ is defined as the unique meromorphic differential two-form with exactly one pole, which is a double pole at $p=q$ with no residue, and with vanishing integral over A_I-cycles $\oint_{A_I} B(p,q) = 0$ (in a canonical basis of cycles (A_I, B^I)). For curves of genus zero as arise from the one-cut solution, the Bergman kernel takes the form

$$B(p,q) \;=\; \frac{dp\,dq}{(p-q)^2}\,. \tag{39}$$

A closely related quantity is a 1-form

$$dE_q(p) \;=\; \frac{1}{2}\int_q^{\bar{q}} B(\xi,p)\,.$$

defined in a neighborhood of a branch point q_i^* in terms of which we may express the recursion kernel

$$K(q,p) \;=\; \frac{dE_q(p)}{\omega(q)}\,. \tag{40}$$

With the above ingredients, we can now present the topological recursion. This recursion determines higher-degree meromorphic differentials $W_n^{(g)}(p_1,\dots,p_n)$ from those of lower degree. The initial data for the recursion are the following one- and two-point correlators of genus zero

$$W_1^{(0)}(p) = 0, \qquad W_2^{(0)}(p_1,p_2) = B(p_1,p_2), \tag{41}$$

with $W_n^{(g)} = 0$, for $g < 0$, by definition. For two sets of indices J and $N = \{1, \dots, n\}$, let us now respectively denote $\vec{p}_J = \{p_i\}_{i\in J}$, and $\vec{p}_N = \{p_1, \dots, p_n\}$. The multi-linear correlators are then defined recursively as

$$\begin{aligned} W_{n+1}^{(g)}(p, \vec{p}_N) = &\sum_{q_i^*} \operatorname*{Res}_{q\to q_i^*} K(q,p)\Big(W_{n+2}^{(g-1)}(q,\bar{q},\vec{p}_N) + \\ &+ \sum_{m=0}^{g}\sum_{J\subset N} W_{|J|+1}^{(m)}(q,\vec{p}_J) W_{n-|J|+1}^{(g-m)}(\bar{q},\vec{p}_{N/J})\Big), \end{aligned} \tag{42}$$

where $\sum_{J\subset N}$ denotes a sum over all subsets J of N. In particular, the lowest order correlators determined by the recursion are

$$W_1^1(p) = \sum_{q_i^*} \operatorname*{Res}_{q\to q_i^*} K(q,p) W_2^0(q,\bar{q})$$

and then [1, 17]

$$W_3^0(p,p_1,p_2) = \sum_{q_i^*} \operatorname*{Res}_{q\to q_i^*} K(q,p)\Big(W_2^0(q,p_1)W_2^0(\bar{q},p_2) + W_2^0(\bar{q},p_1)W_2^0(q,p_2)\Big).$$

Finally, we can define the free energies F_g in genus g. For $g \geq 2$, they come from the corresponding $W_1^{(g)}$:

$$F_g \;=\; \frac{1}{2g-2}\sum_{q_i^*} \operatorname*{Res}_{q\to q_i^*} \phi(q) W_1^{(g)}(q), \tag{43}$$

where $\phi(q) = \int^q y(p)dx(p)$. The expressions F_0 and F_1, given in (5) and (6), can be found independently as discussed, e.g., in [14, 21].

Example 4.1. Let us apply the formalism of the topological recursion to Example 3.1, i.e., the algebraic curve $y^2 = \frac{1}{4}(x-a)(x-b)$ with a parametrization given in (35). We find two branch points $p^* = \pm 1$ and a global expression for the conjugate point (valid around both of these branch points) $\bar{p} = p^{-1}$. As the Gaussian curve represents the one-cut solution on the genus zero spectral curve, the relevant Bergman kernel is simply given by (39), and the recursion kernel (40) takes the form

$$K(q,p) = \frac{8q^3}{(a-b)^2(q^2-1)(p-q)(qp-1)}. \tag{44}$$

Computing the correlators, we find for instance

$$\begin{aligned} W_1^{(1)}(p) \;&=\; -\frac{16p^3\,dp}{(a-b)^2(p^2-1)^4}, \\ W_3^{(0)}(p_1,p_2,p_3) \;&=\; \frac{8dp_1dp_2dp_3}{(a-b)^2}\big((p_1+1)^{-2}(p_2+1)^{-2}(p_3+1)^{-2} \\ &\quad - (p_1-1)^{-2}(p_2-1)^{-2}(p_3-1)^{-2}\big), \\ W_1^{(2)} \;&=\; -\frac{86016dp}{(a-b)^6}\frac{p^7+3p^9+p^{11}}{(p^2-1)^{10}}. \end{aligned} \tag{45}$$

Finally, the genus-two free energy following from the definition (43) reads

$$F_2 = -\frac{16}{15(a-b)^4}. \tag{46}$$

Substituting $a, b = \mp 2\sqrt{T}$ according to (33) we find

$$F_2 = -\frac{1}{240T^2}. \tag{47}$$

Computing higher order free energies, we find [24] well-known expressions for the Euler characteristics [26, 35] of moduli spaces of closed Riemann surfaces of genus g generalizing the above result for genus two, namely,

$$F_g = \frac{B_{2g}}{2g(2g-2)} \frac{1}{T^{2g-2}}. \tag{48}$$

5. Solution of the model

In this section, we derive a formal large N expansion and find free energies F_g of the matrix model (20)

$$Z = \int DH \; e^{-\frac{1}{\hbar}\mathrm{tr}V(H)} = \; \exp \sum_{g=0}^{\infty} \hbar^{2g-2} F_g, \tag{49}$$

with the potential given in (19), i.e., $V(H) = \frac{1}{2}H^2 - \frac{stH}{1-tH}$. The potential in this model is a function of parameter s which generates the number of backbones and t which generates the number of chords. Thus, the F_g are also functions of these two parameters. Expanding F_g as a series in s and t we obtain multiplicities labeled by the genus, number of backbones and number of chords, which provide the solution of our model. In fact, free energies depend also on 't Hooft coupling $T = \hbar N$, however, this dependence is rather simple, and all the information we seek can be obtained by setting $T = 1$. To solve this matrix model, we follow the steps explained in previous sections: we find a parametrization of the spectral curve and then apply the formalism of the topological recursion to derive associated multi-linear correlators $W_n^{(g)}(p_1, \ldots, p_n)$ and free energies F_g. Finally, we present explicit expressions for free energies for small values of g and show that they indeed encode the $c_{g,b}(n)$ discussed in the introduction, i.e., the numbers of chord diagrams for various small genera and numbers of backbones.

5.1. **Spectral curve.** We begin by determining a specific spectral curve. Let us note that the potential (19) is in fact not bounded below, and there is a singularity as $x \to t^{-1}$. Nonetheless, this singularity does not pose a problem in the formal expansion of the free energy as we will see. Also for this reason, we treat the parameters s and t in the potential (19) as generating parameters for powers of x, and so we effectively consider the behavior of eigenvalues only in the vicinity of the perturbed Gaussian minimum even though there are three critical points of the potential, i.e., we are studying

the one-cut solution of the model. We can therefore determine the one-cut semiclassical resolvent $\omega_1^{(0)}(x)$ using the Migdal integral (27) leading to

$$\begin{aligned}\omega_1^{(0)}(x) \;=\; & \frac{x(xt-1)^2-st}{2T(xt-1)^2}+\frac{\sqrt{(x-a)(x-b)}}{2T}\times \\ & \times\Big(\frac{st^2\big(4+(xt^2-3t)(a+b)-2xt+2abt^2\big)}{2(xt-1)^2\big((at-1)(bt-1)\big)^{3/2}}-1\Big),\end{aligned} \tag{50}$$

where a and b denote cut endpoints. Expanding this expression for large x we find that it has a form $\omega_1^{(0)}(x)=c_1+c_2x^{-1}+\mathcal{O}(x^{-2})$. However, imposing the asymptotics given in (23) requires that $c_1=0$ and $c_2=1$. These two equations take the following form

$$\begin{cases} 0=a+b+\frac{st(at+bt-2)}{\big((at-1)(bt-1)\big)^{3/2}}, \\ 16T=(a-b)^2+\frac{4s\big((2-\frac{(a+b)t}{2})(at+bt-2)+2abt^2-3t(a+b)+4\big)}{\big((at-1)(bt-1)\big)^{3/2}}. \end{cases} \tag{51}$$

In principle, one could now solve (51) for a and b and substitute these values back into (50) to get the exact formula for the resolvent, but we know of no closed-form solution. However, one can determine perturbative expansion of a and b in s and t, and such a representation is sufficient at least as far as perturbative expansion of free energies is concerned. For example, we find that the perturbative expansions in s (with exact dependence on t) start as

$$\begin{cases} a=a(s,t)=-2T^{1/2}+st\frac{(1-4t^2T)^{-1/2}}{1+2tT^{1/2}}+\dots, \\ b=b(s,t)=2T^{1/2}+st\frac{(1-4t^2T)^{-1/2}}{1-2tT^{1/2}}+\dots \end{cases} \tag{52}$$

and an analogous expansion in t (with exact dependence on s) can be determined.

It is convenient to introduce half the sum and difference of the cut endpoints a,b

$$S=\frac{a+b}{2}, \qquad D=\frac{a-b}{2}. \tag{53}$$

Taking the square of the first equation in (51) as well as determining the denominator $\big((at-1)(bt-1)\big)^{-3/2}$ from both equations in (51) leads to the following pair of equations

$$\begin{cases} 4s^2t^4(3D^2-4T)^6=D^2(D^2-4T)^2\big(4D^2-t^2(3D^2-4T)^2\big)^3, \\ S=\frac{D^2-4T}{t(3D^2-4T)}. \end{cases} \tag{54}$$

The first equation is a polynomial equation for D which can be solved perturbatively to any accuracy in s or t, and then the second equation immediately determines S. In appendix B, we present a perturbative expansion of S and D in s with exact dependence on t; from this one can easily find higher order terms in the expansion (52).

Theorem 5.1. *Provided the 't Hooft parameter $T \neq 0$, there is a unique continuous extension of the Gaussian to a solution of (54). In fact, the extension is smooth in t and converges for all s and small t, and it furthermore satisfies $a(s,-t) = -b(s,t)$.*

Proof. Consider the polynomial $q_{s,t}(S,D)$ given by the difference of the left- from right-hand side of the first equation (54). Direct computation shows that $\frac{\partial q_{s,t}(S,D)}{\partial D}\Big|_{t=0} = 512\ D^{15}(8T-3D^4)(4T-D^4)$. Provided this is non-vanishing, the Implicit Function Theorem applies to provide the unique smooth and convergent one-cut solution. Meanwhile for the Gaussian with $t=0$, we have cut endpoints $\pm 2\sqrt{T}$, so $S=0$ and $D=2\sqrt{T}$. Thus, $T \neq 0$ implies $D \neq 0$, and $D^4 = 16T$ further implies that the factors $(8t-3D^4) = -8T$ and $(4T-D^4) = -12T$ are also non-zero as required.

The asserted symmetry $a(s,t) = -b(s,-t)$ is equivalent to the statement that S is an odd and D an even function of t. Meanwhile, the first equation in (54) involves only even powers of t so the smooth solution from the Implicit Function Theorem must be an even function of t. Insofar as T is independent of t, it is even as well, so the second equation in (54) displays S as an odd function of t. □

In what follows, we will work with general expressions for the resolvent and ultimately derive exact expressions for F_g in terms of S and D. Substituting the above expansions, we will then find perturbative expansions of F_g in s and t. Let us therefore proceed to find the spectral curve in terms of parameters a and b or equivalently S and D. This curve is to be expressed in terms of variables x and y defined as in (28). In particular, only expressions involving square roots $\sqrt{(x-a)(x-b)}$, given in the second line of (50), will enter the formula for y. Moreover, note that the second line of (50) involves one non-rational expression $\big((at-1)(bt-1)\big)^{3/2}$, which however can be replaced by a rational function using the first equation in (51). In consequence, the equation for the spectral curve takes form

$$y = \frac{1}{2}\sqrt{(x-a)(x-b)}\frac{\big(tx-1+\frac{(a+b)t}{4}\big)^2+\gamma}{(tx-1)^2}, \tag{55}$$

so that comparing with (30) we find

$$M(x) = \frac{\big(tx-1+\frac{(a+b)t}{4}\big)^2+\gamma}{2(tx-1)^2}. \tag{56}$$

The entire dependence in (55) on x and y is given explicitly, and γ is a rational function of a, b and t given below. If we square this equation we obtain an algebraic equation for the spectral curve. In fact, notice that all factors of t in this equation multiply a and b, and if we rescale also x and y by t, then the entire dependence on t can be encoded in rescaled variables

$$\xi = xt, \quad \eta = yt, \quad \alpha = at, \quad \beta = bt, \quad \sigma = St, \quad \delta = Dt. \tag{57}$$

In particular in terms of variables introduced above, the expression for γ takes the form

$$\begin{aligned}\gamma &= -\frac{(\alpha+\beta)(\alpha^2+\beta^2+14(\alpha+\beta-\alpha\beta)-16)}{16(\alpha+\beta-2)}\\ &= \frac{\sigma(4-4\delta^2-7\sigma+3\sigma^2)}{4(\sigma-1)}.\end{aligned} \tag{58}$$

Finally in these rescaled variables, the remaining equation reads

$$(\xi-1)^4\eta^2=\frac{1}{4}(\xi-\alpha)(\xi-\beta)\Big[(\xi-1+\frac{\alpha+\beta}{4})^2+\gamma\Big]^2. \tag{59}$$

This is our desired expression for the spectral curve. In order to apply the topological recursion, we must further find a parametric form of this equation. To this end, note that the square root of the above equation, or equivalently the equation (55), is a modification of the Gaussian curve $y=\frac{1}{2}\sqrt{(x-a)(x-b)}$ given in (34) by a rational function in x. The essence of the parametrization of the Gaussian curve given in (35) was to take care of this square root and devise a function $x=x(p)$ such that $\frac{1}{4}(x-a)(x-b)$ is a complete square of a rational function in p, which is then identified with y. We can therefore borrow this aspect of the Gaussian parametrization (35) and just modify the form of y by multiplying it by this additional rational function of x in the present case. This leads to the following parametric form of the spectral curve

$$\begin{cases}\xi(p)=\frac{\alpha+\beta}{2}+\frac{\alpha-\beta}{4}\Big(p+\frac{1}{p}\Big)=\sigma+\frac{\delta}{2}\Big(p+\frac{1}{p}\Big),\\ \eta(p)=\frac{\alpha-\beta}{8}\Big(\frac{1}{p}-p\Big)\frac{\left(\xi-1+\frac{\alpha+\beta}{4}\right)^2+\gamma}{(\xi-1)^2}=\frac{\delta}{4}\Big(\frac{1}{p}-p\Big)\frac{\left(\xi-1+\frac{\sigma}{2}\right)^2+\gamma}{(\xi-1)^2}.\end{cases} \tag{60}$$

With this parametrization of the spectral curve, we are prepared to compute the free energies using the topological recursion.

5.2. **Free energies.** We can now determine free energies F_g, for genus $g\geq 2$, of the spectral curve given in (59) as parametrized by (60). This curve has genus zero, so the Bergman kernel takes the form (39). As the parametrization of the ξ variable is the same as in the Gaussian case (35), we find the same branch points $p=\pm 1$ and globally conjugate points $\bar{p}=p^{-1}$ as given in Example 3.1. Nonetheless, all other quantities which depend also on η have of course a much more complicated form than in the Gaussian case. In particular, the recursion kernel takes the form

$$K(q,p)=\frac{2q^3}{(q^2-1)(p-q)(qp-1)\delta^2}\,\frac{\left((q^{-1}+q)\delta+2\sigma-2\right)^2}{4\gamma+\left((q^{-1}+q)\delta+3\sigma-2\right)^2}. \tag{61}$$

With these ingredients, we can compute multi-linear correlators $W_n^{(g)}$. They are given by rather complicated and not very enlightening formulas, which

are too lengthy to write down here. However, just to present the simplest example, let us give the formula for $W_3^{(0)}$:

$$
\begin{aligned}
W_3^{(0)}(p_1,p_2,p_3) = & 2dp_1dp_2dp_3\,\frac{\sigma-1}{\sigma^2}\times \\
& \Big(\frac{\sigma-\delta-1}{(p_1+1)^2(p_2+1)^2(p_3+1)^2(1+\delta-4\sigma+3\sigma^2)} \\
& -\frac{\sigma+\delta-1}{(p_1-1)^2(p_2-1)^2(p_3-1)^2(1-\delta-4\sigma+3\sigma^2)}\Big).
\end{aligned} \tag{62}
$$

At this stage viz. Theorem 5.1, we can also confirm that all our results are consistent with the Gaussian model discussed in Examples 3.1 and 4.1. The potential (19) reduces to the Gaussian potential when $st = 0$. In both cases $s = 0$ and $t = 0$, the system (51) simplifies so that the first equation implies $a = -b$, and then the second equation leads to the result (33) also in agreement with the expansion (52). For $a = -b$, we also find that $\gamma = 0$, and the curve (59) reduces to the Gaussian curve (34). Furthermore, the recursion kernel (61) and all correlators and (62) in particular reduce respectively to (44) and (45).

Finally with Mathematica [43], we compute genus two and genus three free energies. As these are the free energies of the curve (60), which involves variables x and y rescaled by t, let us denote them by $\widetilde{F}_2$ and $\widetilde{F}_3$. We find they are given by the following exact expressions

$$
\begin{aligned}
\widetilde{F}_2 = & -\frac{(1-\sigma)^2}{240\delta^4(1-\delta-4\sigma+3\sigma^2)^5(1+\delta-4\sigma+3\sigma^2)^5}\times \\
& \times\Big(160\delta^4(1-3\sigma)^4(1-\sigma)^6-80\delta^2(1-3\sigma)^6(1-\sigma)^8 \\
& +16(1-3\sigma)^8(1-\sigma)^{10}+\delta^{10}(-16+219\sigma-462\sigma^2+252\sigma^3) \\
& +10\delta^6(1-3\sigma)^2(1-\sigma)^4(-16-126\sigma-423\sigma^2+2286\sigma^3-2862\sigma^4+1134\sigma^5) \\
& +5\delta^8(1-\sigma)^2(16+189\sigma-2970\sigma^2+9549\sigma^3-11286\sigma^4+4536\sigma^5)\Big)
\end{aligned}
$$

$$
\begin{aligned}
\widetilde{F}_3 = {} & \frac{(1-\sigma)^4}{2016\delta^8(1-\delta-4\sigma+3\sigma^2)^{10}(1+\delta-4\sigma+3\sigma^2)^{10}}\times \\
& \times\Big(107520\delta^8(1-3\sigma)^8(1-\sigma)^{12}-61440\delta^6(1-3\sigma)^{10}(1-\sigma)^{14} \\
& +23040\delta^4(1-3\sigma)^{12}(1-\sigma)^{16}-5120\delta^2(1-3\sigma)^{14}(1-\sigma)^{18}+512(1-3\sigma)^{16}(-1+\sigma)^{20} \\
& +\delta^{20}(512-48603\sigma+471708\sigma^2-1645251\sigma^3+2623320\sigma^4-1957284\sigma^5+555660\sigma^6) \\
& +\delta^{18}(1-\sigma)^2(-5120-639867\sigma+20457090\sigma^2-177935466\sigma^3+697802346\sigma^4 \\
& -1439492067\sigma^5+1623708828\sigma^6-948940272\sigma^7+225042300\sigma^8) \\
& +252\delta^{10}(1-3\sigma)^6(1-\sigma)^{10}(-512+2970\sigma+70785\sigma^2-3780\sigma^3-1030050\sigma^4 \\
& +2252880\sigma^5-1077300\sigma^6-1682100\sigma^7+2417850\sigma^8-1129950\sigma^9+178605\sigma^{10}) \\
& +84\delta^{12}(1-3\sigma)^4(1-\sigma)^8(1280-17523\sigma+256203\sigma^2+5113206\sigma^3-22143348\sigma^4+2780703\sigma^5 \\
& +124222086\sigma^6-273630879\sigma^7+265968360\sigma^8-126660105\sigma^9+24111675\sigma^{10}) \\
& +9\delta^{16}(1-\sigma)^4(2560+156387\sigma+605388\sigma^2-68079487\sigma^3+696887576\sigma^4-3203644311\sigma^5 \\
& +8142105636\sigma^6-12207659661\sigma^7+10771521264\sigma^8-5182076088\sigma^9+1050197400\sigma^{10}) \\
& +3\delta^{14}(1-3\sigma)^2(1-\sigma)^6(-20480-6699\sigma-23380056\sigma^2+166860351\sigma^3 \\
& +520822512\sigma^4-6709872897\sigma^5+22111131168\sigma^6-36421137675\sigma^7 \\
& +33054947856\sigma^8-15849836352\sigma^9+3150592200\sigma^{10})\Big)
\end{aligned}
$$

The free energies F_g in our model are closely related to these results. Indeed, the original curve is written in terms of x and y while the above free energies are computed for the curve in rescaled variables (57). In general, rescaling a variable x or y by t results in rescaling of the genus g free energy by t^{2g-2}. Because we rescale both x and y, this means that the free energies are rescaled by $t^{2(2g-2)}$. The genus two and three free energies, with dependence on t reintroduced, are thus given as

$$F_2 = t^4\widetilde{F}_2, \qquad F_3 = t^8\widetilde{F}_3. \tag{63}$$

Note that the entire dependence on t arises upon substituting σ and δ according to (57).

The expressions (63) are our desired results, and all we need do is substitute the expressions for σ and δ or in fact a and b given in principle by solving the equations (51). In practice however, what we can do is to substitute the series expansion of a and b in terms of s and t given in (52) or to better accuracy in appendix B. This finally leads to the ultimate expansion of F_2 and F_3 in s and t which we seek. Let us present results after setting

$$T = 1 \text{ and } z = t^2. \tag{64}$$

We can consider the perturbative expansion of F_g in s with exact dependence on t or the perturbative expansion in t with exact dependence on s. In the

former case, we find

$$
\begin{aligned}
F_2(s,t) &= -\frac{1}{240} + \sum_{b\geq 1} \frac{s^b}{b!} C_{2,b}(t^2), \\
F_3(s,t) &= \frac{1}{1008} + \sum_{b\geq 1} \frac{s^b}{b!} C_{3,b}(t^2),
\end{aligned}
\tag{65}
$$

with the constant terms $-1/240$ and $1/1008$ reproducing the free energies of the Gaussian model (48) in accordance with the fact that our model reduces to the Gaussian one for $st = 0$.

We find that indeed $C_{2,1}(z) \equiv C_2(z)$ and $C_{3,1}(z) \equiv C_3(z)$ reproduce $C_g(z)$ given in (12), and $C_{2,b}(z)$ and $C_{3,b}(z)$, for $b = 2, 3$, reproduce generating functions given in (18). This confirms that our model correctly reproduces these known results for $b = 1, 2, 3$ backbones in genera two and three. Moreover, we can easily perform expansions to higher order in s revealing generating functions for more backbones. For example for four backbones, we find

$$
\begin{aligned}
C_{2,4}(z) &= \frac{144z^7}{(1-4z)^{13}}(38675 + 620648z + 2087808z^2 \\
&\quad + 1569328z^3 + 134208z^4), \\
C_{3,4}(z) &= \frac{48z^9}{(1-4z)^{16}}(53416125 + 1194366915z + 6557325096z^2 \\
&\quad + 10738411392z^3 + 4580024832z^4 + 236239616z^5),
\end{aligned}
$$

and for five backbones, we find

$$
\begin{aligned}
C_{2,5}(z) &= \frac{144z^8}{(1-4z)^{31/2}}(2543625 + 62424520z + 375044396z^2 \\
&\quad + 671666053z^3 + 314761848z^4 + 18335696z^5), \\
C_{3,5}(z) &= \frac{720z^{10}}{(1-4z)^{37/2}}(360380790 + 11275076865z + 95744892585z^2 \\
&\quad + 282797424880z^3 + 291167707410z^4 + 85497242928z^5 + 3218434848z^6).
\end{aligned}
$$

We can also determine the perturbative expansion in t with exact dependence on s. Again scaling variables as in (64) and now expanding up to seven chords, i.e., to order z^7, we find

$$
\begin{aligned}
F_2 &= -\frac{1}{240} + 21sz^4 + (483s + 1485\frac{s^2}{2})z^5 + (6468s + 26808s^2 + 15015s^3)z^6 \\
&\quad + (66066s + 526064s^2 + 768564s^3 + 232050s^4)z^7 + \mathcal{O}(z^8), \\
F_3 &= \frac{1}{1008} + 1485sz^6 + (56628s + 225225\frac{s^2}{2})z^7 + \mathcal{O}(z^8),
\end{aligned}
$$

These expansions provide the numbers of configurations with a fixed number of chords and arbitrary numbers of backbones in accordance with Proposition 1.1. For example, for genus two with seven chords, there are at most four backbones, and for genus three with seven chords, there are at most two backbones. We can easily continue the above expansions to higher powers of s and t.

Appendix A. Genera zero and one free energies

In this appendix, we determine free energies for genus zero and one, i.e., compute F_0 and F_1 in general in notation introduced in section 3, and ultimately substitute to derive results for our model in section 5.

A.1. **Genus zero.** The genus zero free energy for the matrix model given in (20) can be written in a saddle point approximation [12, 29] as

$$\begin{aligned} F_0 &= -T\int dx\rho(x)V(x) + T^2\int dxdy\rho(x)\rho(y)\log|x-y| + \\ &\quad +T\lambda\Big(\int \rho(x)dx - t_0\Big) \\ &= -\frac{T}{2}\Big(\int dx\rho(x)V(x) - \lambda t_0\Big). \end{aligned} \tag{66}$$

In this variational problem, it is crucial to impose the condition preserving the number of eigenvalues, which is fixed by (24). This can be done by introducing and subsequently determining the Lagrange multiplier λ, where the rest of the notation in the above equation is the same as in section 3. The Lagrange multiplier can be determined from the asymptotic condition on $y(x)$ which (for the one-cut solution) leads to the formula

$$\lambda = \lim_{\Lambda}\Big(2\int_b^{\Lambda} y(x)dx\ - V(\Lambda) - t_0\log\Lambda\Big), \tag{67}$$

which is finite as singular terms from the upper integration limit must be canceled by $V(\Lambda)$ and $t_0\log\Lambda$ in the limit of large Λ. Substituting the results found in section 5, we find

$$\lambda = \frac{tD^2 + S(3t^2D^2 + 2tS(5-3tS) - 4) + 2tD^2(3tS-1)\log(-D/2)}{8t(tS-1)}, \tag{68}$$

and t_0 defined in (24) takes form

$$t_0 = \frac{(a-b)^2(3(a+b)t-2)}{16T\big((a+b)t-2\big)} = \frac{D^2(3tS-1)}{4T(tS-1)}. \tag{69}$$

Moreover, we find the following integral

$$\int dx\rho(x)V(x) = -\frac{1}{64t^3T(tS-1)}\Big(f_1 + \frac{f_2}{\sqrt{(tS+tD-1)(tS-tD-1)}}\Big), \tag{70}$$

where

$$\begin{aligned} f_1 &= 16S - 16(s+3S^2)t + 8S(-3D^2+2s+6S^2)t^2 + \\ &\quad + (D^4 - 16S^4 + 4D^2(2s+7S^2))t^3 + 3D^2S(D^2 - 4(2s+S^2))t^4 \\ f_2 &= 8(-2S + 2(s+4S^2)t - 4S(-D^2+s+3S^2)t^2 + \\ &\quad + 2(S^2-D^2)(s+4S^2)t^3 + S(D^2(3s+4S^2) - 2D^4 - 2S^4)t^4). \end{aligned}$$

Substituting now (68), (69) and (70) into (66), we find F_0 expressed in terms of cut endpoints a and b or rather their combinations S and D defined in (53). We can now substitute expansions of S and D obtained from solving (51) or (54). In particular, substituting perturbative expansions of S and D in s with exact dependence on t as given in appendix B, setting $T = 1$ and $z = t^2$ as in (64), and neglecting z-independent contributions which reproduce the genus zero result in the Gaussian case, we find perfect agreement with known results.

Indeed, $C_{0,1}(z) \equiv C_0(z)$ reproduces the generating function for the Catalan numbers given in (10) while $C_{0,2}(z)$ and $C_{0,3}(z)$ agree with (18). Moreover, we can perform expansions in s to arbitrary order and find exact generating functions for arbitrary numbers of backbones. In particular for four and five backbones, we find

$$\begin{aligned} C_{0,4}(z) &= \frac{24z^3(3+18z+8z^2)}{(1-4z)^7} = \\ &= 72z^3 + 2448z^4 + 44544z^5 + 585984z^6 + \dots \qquad (71) \\ C_{0,5}(z) &= \frac{24z^4(55+741z+1512z^2+336z^3)}{(1-4z)^{19/2}} = \\ &= 1320z^4 + 67944z^5 + 1765440z^6 + \dots \qquad (72) \end{aligned}$$

To test these results and to provide an idea of their combinatorial complexity, we show in appendix C by an explicit enumeration that in genus zero on four backbones there are indeed 72 diagrams with three chords and 2448 diagrams with four chords.

A.2. **Genus one.** As shown in [14], the genus one free energy is given by (6), where $M(x)$ is in general defined via (30) and in our model takes the form (56), where a and b are given by a solution to the system (51). We can expand this result to arbitrary accuracy in s or t to find generating functions for chord diagrams with arbitrary numbers of backbones or chords. The constant term is logarithmic in T and merely reproduces the Gaussian genus one free energy, Setting $T = 1$ in accordance with (64), expanding in s, keeping the exact dependence in t and using results presented in appendix B, we find that indeed $C_{1,1}(z) \equiv C_1(z)$ agrees with $C_1(z)$ given in (12) and $C_{1,2}(z)$, $C_{1,3}(z)$ agree with (18). This confirms that our approach correctly reproduces known results in genus one for three or fewer backbones. Moreover, performing expansions to higher order in s, we can immediately find

generating functions for more backbones, for instance,

$$\begin{aligned} C_{1,4}(z) &= \frac{24z^5(715+7551z+12456z^2+2096z^3)}{(1-4z)^{10}}, \\ C_{1,5}(z) &= \frac{48z^6(13650+253625z+975915z^2+840600z^3+86448z^4)}{(1-4z)^{25/2}}. \end{aligned}$$

Appendix B. Cut endpoints

It is convenient to present here a solution for cut endpoints a and b obtained from equations (51) in terms of variables introduced in (53)

$$S = \frac{a+b}{2}, \qquad D = \frac{a-b}{2}. \tag{73}$$

Equations (54) provide a solution which is perturbative in s and exact in t:

$$\begin{aligned} D = &-2T^{1/2} - \frac{2st^2T^{1/2}}{(1-4t^2T)^{3/2}} - \frac{3s^2t^4T^{1/2}(3+4t^2T)}{(1-4t^2T)^4} \\ &- \frac{s^3t^6T^{1/2}(47+308t^2T+128t^4T^2)}{(1-4t^2T)^{13/2}} \\ &- \frac{s^4t^8T^{1/2}(1051+15756t^2T+33168t^4T^2+6720t^6T^3)}{4(1-4t^2T)^9} \\ &- \frac{3s^5t^{10}T^{1/2}(2037+53448t^2T+262224t^4T^2+272640t^6T^3+32768t^8T^4)}{4(1-4t^2T)^{23/2}} + \mathcal{O}(s^6) \\ S = &\frac{st}{(1-4t^2T)^{3/2}} + \frac{2s^2t^3(1+8t^2T)}{(1-4t^2T)^4} \\ &+ \frac{s^3t^5(7+166t^2T+280t^4T^2)}{(1-4t^2T)^{13/2}} \\ &+ \frac{2s^4t^7(15+684t^2T+3552t^4T^2+2560t^6T^3)}{(1-4t^2T)^9} \\ &+ \frac{s^5t^9(143+10408t^2T+107094t^4T^2+240144t^6T^3+96096t^8T^4)}{(1-4t^2T)^{23/2}} + \mathcal{O}(s^6) \end{aligned}$$

Appendix C. Explicit diagram counting

To illustrate the combinatorial complexity of chord diagrams and fully appreciate the generating functions which we have computed, we directly count in this appendix the number of genus zero diagrams on four backbones with three or four chords. In this way, we also confirm correctness of the coefficients $c_{0,4}(3) = 72$ and $c_{0,4}(4) = 2448$ in the genus zero expansion on four backbones in (7).

Our first task is to distribute endpoints of chords along backbones in all possible inequivalent ways. If there are n chords, then there are $2n$ such endpoints, perhaps better apprehended as half-chords. We shall denote a

partition of $\{1, 2, \cdots 2n\}$ into n_i sets of cardinality $i \geq 1$ as $\{n_1, \ldots, n_K\}$ where $\sum_{k=1}^{K} n_k = 2n$.

We must then count how many ways there are to produce a connected chord diagram of the correct genus by pairing up these putative endpoints. In fact in the cases under consideration, the genus is always zero by Proposition 1.1, so here we must count simply the number of connected such pairings.

To begin, consider diagrams on four backbones with three chords. The partitions of six are: $\{3, 1, 1, 1\}$ and $\{2, 2, 1, 1\}$.

- For $\{3, 1, 1, 1\}$, all possible linearly ordered sets of backbones arise by embedding them in the real line with the induced linear order and orientations. To achieve such an embedding, we merely determine which among the four backbones has three half-chords. Order these half-chords 1, 2, 3, which connect uniquely to the other backbones occurring in the order 1, 2, 3 along the real axis. There are thus 4 times $3! = 24$ such chord diagrams.
- For $\{2, 2, 1, 1\}$, there are $\binom{4}{2} = 6$ places to put the two 2's. By connectivity, there must be a chord between the two 2's, and there are four ways to do this. The two remaining half-chords on the 2's again come in an order along the real axis as do the two remaining 1's, giving another factor $2!$, for a total of $6 \cdot 4 \cdot 2! = 48$.

The grand total is thus $24 + 48 = 72$ diagrams in agreement with the first term in the expansion of the generating function $C_{0,4}$ given in (7).

We next consider the more involved case of four backbones with four chords. There are the five partitions $\{5, 1, 1, 1\}$, $\{4, 2, 1, 1\}$, $\{3, 2, 2, 1\}$, $\{2, 2, 2, 2\}$, and $\{3, 3, 1, 1\}$ of the eight half-chords into four backbones, and:

- For $\{5, 1, 1, 1\}$, there are four locations for the 5. Two half-chords on 5 must be connected, and there are $\binom{5}{2} = 10$ of these. The remaining half-chords on 5 connect arbitrarily to the 1's giving another factor $3!$, for a total of $4 \cdot 10 \cdot 6 = 240$ diagrams for $\{5, 1, 1, 1\}$.
- For $\{4, 2, 1, 1\}$ there are four locations for 4 and three for 2. There must be at least one 4-2 chord by connectivity and at most two by cardinality.

  If there is a unique chord 4-2, among the $4 \cdot 2$ such possibilities, then the other half-chord on the 2 must connect to a 1 by definition, and there are two of these; the remaining 1 then can connect to any of the three remaining half-chords; there are a total of $4 \cdot 3 \cdot 4 \cdot 2 \cdot 2 \cdot 3 = 576$ in this sub-case.

  There are $\binom{4}{2} = 6$ times two possible pairs of 4-2 chords and two ways to produce a connected diagram connecting the remaining 1's to the remaining half-chords on 4; there are a total of $4 \cdot 3 \cdot 6 \cdot 2 \cdot 2 = 288$ in this sub-case.

There are a total of $576 + 288 = 864$ chord diagrams for $\{4, 2, 1, 1\}$.

- For $\{3, 2, 2, 1\}$, there are four locations for 3 and three for 1. There must be at least one 3-2 chord by connectivity.

  If there is exactly one 3-2 chord among the $3 \cdot 4$ possibilities, then the other half-chord from the 2 must go to the other 2, with two possibilities. This gives $4 \cdot 3 \cdot 4 \cdot 3 \cdot 2 = 288$ in this sub-case.

  If there are exactly two 3-2 chords, then their 2-endpoints must lie on different backbones by connectivity. There are thus $\binom{3}{2} = 3$ times eight such pairs of chords and a unique completion to a diagram with exactly two 3-2 chords. There are a total of $4 \cdot 3 \cdot 3 \cdot 8 = 288$ in this sub-case.

  There are $4 \cdot 3 \cdot 2$ configurations of three 3-2 chords with a unique completion, so a total of $4 \cdot 3 \cdot 4 \cdot 3 \cdot 2 = 288$ in the sub-case of exactly three 3-2 chords.

  There are a total of $288 + 288 + 288 = 864$ diagrams for $\{3, 2, 2, 1\}$.

- For $\{2, 2, 2, 2\}$, there can be no two chords with the same backbone endpoints by connectivity, so the first chord on the first backbone lands in one of six places and the second in one of four places on different backbones. There are two connected completions of any such partial diagram, for a total of $6 \cdot 4 \cdot 2 = 48$ diagrams for $\{2, 2, 2, 2\}$.

- For $\{3, 3, 1, 1\}$, there are $\binom{4}{2}$ locations for the 3's. A fixed 1 must connect to a 3 in one of three possible ways times two for the two 3's. Either this 3 connects to the other 1 in two possible ways and to the other 3 in three possible ways, or not, in which case, the other 1 connects to the other 3 in three possible ways with two possible connected completions for a total of $6 \cdot 6 \cdot (2 \cdot 3 + 3 \cdot 2) = 432$ diagrams for $\{3, 3, 1, 1\}$.

Overall, we find $240 + 864 + 864 + 48 + 432 = 2448$ chord diagrams in agreement with the second term $c_{0,4}(4)$ in the expansion of the generating function $C_{0,4}$ given in (7).

Acknowledgements: JEA and RCP are supported by the Centre for Quantum Geometry of Moduli Spaces which is funded by the Danish National Research Foundation. The research of LCh is supported by the Russian Foundation for Basic Research (Grants Nos 10-02-01315-a, 11-01-12037-ofi-m-2011, and 11-02-90453-Ukr-f-a) by the Program Mathematical Methods of Nonlinear Dynamics and by the Grant of Supporting Leading Scientific Schools NSh-4612.2012.1. The research of PS is supported by the DOE grant DE-FG03-92-ER40701FG-02, the European Commission under the Marie-Curie International Outgoing Fellowship Programme, and the Foundation for Polish Science. RCP also acknowledges the kind support of Institut Henri Poincaré where parts of this manuscript were written.

References

[1] A. Alexandrov, A. Mironov, A. Morozov, *Partition functions of matrix models as the first special functions of string theory: Finite Hermitian one-matrix model*, Theor. Math. Phys. 142 (2005) 349-411.

[2] J. Ambjorn, L. Chekhov, C.F. Kristjansen, Yu. Makeenko, *Matrix Model Calculations beyond the Spherical Limit*, Nucl.Phys. B404 (1993) 127-172; Erratum-ibid. B449 (1995) 681.

[3] J. Ambjorn, L. Chekhov, Yu. Makeenko, *Higher Genus Correlators from the Hermitian One-Matrix Model*, Phys. Lett. B282 (1992) 341-348.

[4] J. E. Andersen, F. W. D. Huang, R. C. Penner, C. M. Reidys, *Topology of RNA-RNA interaction structures*, (2112), arXiv:1112.6194.

[5] J. E. Andersen, R. C. Penner, C. M. Reidys, M. S. Waterman, *Topological classification and enumeration of RNA structures by genus*, preprint.

[6] D. Bessis, C. Itzykson, J. B. Zuber, *Quantum field theory techniques in graphical enumeration*, Adv. Appl. Math. **1** (1980), 109–157.

[7] M. Bon, G. Vernizzi, H. Orland, A. Zee, *Topological classification of RNA structures*, Jour. Math. Bio. **379**, (2008), 900-911.

[8] G. Borot, B. Eynard, M. Mulase, B. Safnuk, *A matrix model for simple Hurwitz numbers, and topological recursion*, 0906.1206 [math-ph].

[9] V. Bouchard, A. Klemm, M. Marino, S. Pasquetti, *Remodeling the B-model*, Commun. Math. Phys. 287 (2009) 117-178.

[10] V. Bouchard, A. Catuneanu, O. Marchal, P. Sułkowski, *The remodeling conjecture and the Faber-Pandharipande formula*, 1108.2689 [math.AG].

[11] V. Bouchard, P. Sułkowski, *Topological recursion and mirror curves* , 1105.2052 [hep-th].

[12] E. Brezin, C. Itzykson, G. Parisi, J. Zuber, *Planar diagrams*, Commun. Math. Phys. 59 (1978) 35-51.

[13] K. Chapman, M. Mulase, B. Safnuk, *The Kontsevich constants for the volume of the moduli of curves and topological recursion*, 1009.2055 [math.AG].

[14] L. Chekhov, *Genus one correlation to multi-cut matrix model solutions*, Theor. Math. Phys. 141 (2004) 1640-1653, Teor. Mat. Fiz. 141 (2004) 358-374.

[15] L. Chekhov, B. Eynard, *Hermitean matrix model free energy: Feynman graph technique for all genera*, JHEP 0603 (2006) 014.

[16] L. Chekhov, B. Eynard, N. Orantin, *Free energy topological expansion for the 2-matrix model*, JHEP 0612 (2006) 053.

[17] L. Chekhov, A. Marshakov, A. Mironov, D. Vasiliev, *DV and WDVV*, Phys. Lett. B562 (2003) 323-338.

[18] M. G. dell'Erba, G. R. Zemba, *Thermodynamics of a model for RNA folding*, Phys. Rev. E **79** (2009), 011913.

[19] B. Eynard, *All genus correlation functions for the hermitian 1-matrix model*, JHEP 0411 (2004) 031.

[20] B. Eynard, *Invariants of spectral curves and intersection theory of moduli spaces of complex curves*, 1110.2949 [math-ph].

[21] B. Eynard and N. Orantin, *Invariants of algebraic curves and topological expansion* Comm. Number Th. and Phys. **1** (2007), 347-452.

[22] B. Eynard, N. Orantin, *Algebraic methods in random matrices and enumerative geometry* , 0811.3531 [math-ph].

[23] I. Garg and N. Deo, *RNA matrix model with external interactions and their asymptotic behaviors* Phys. Rev. E **79**, (2009), 061903.

[24] S. Gukov, P. Sułkowski, *A-polynomial, B-model, and Quantization*, JHEP 1202 (2012) 070, [1108.0002 [hep-th]].

[25] S. W. Han and C. M. Reidys, private communication.

[26] J. Harer and D. Zagier, *The Euler characteristic of the moduli space of curves*, Invent. Math. **85** (1986), 457-485.

[27] M. Kontsevich, *Intersection theory on the moduli space of curves and the matrix Airy function*, Comm. Math. Phys. **147** (1992), 1–23.

[28] S. Lando and A. K. Zvonkin, Graphs on Surfaces and their Applications, Springer, New York (2004).

[29] M. Mariño, *Les Houches lectures on matrix models and topological strings*, In: Applications of Random Matrices in Physics (Les Houches Lecture Notes) NATO Sci. Ser. **221** (2005) Springer, New York, 319-378.

[30] A. Morozov and S. Shakirov, *From Brezin-Hikami to Harer-Zagier formulas for Gaussian correlators*, hep-th 1007.4100.

[31] M. Mulase, M. Penkava, *Topological recursion for the Poincare polynomial of the combinatorial moduli space of curves*, 1009.2135 [math.AG].

[32] P. Norbury, *Cell decompositions of moduli space, lattice points and Hurwitz problems*, 1006.1153 [math.GT].

[33] P. Norbury, N. Scott, *Gromov-Witten invariants of* $\mathbb{P}^1$ *and Eynard-Orantin invariants*, 1106.1337 [math.AG].

[34] H. Orland and A. Zee, *RNA folding and large N matrix theory*, Nucl. Phys. B, **620** (2002), 456-476.

[35] R. C. Penner, *Perturbative series and the moduli space of Riemann surfaces,* J. Diff. Geom. **27** (1988), 35-53.

[36] —, *The structure and singularities of quotient arc complexes*, Jour. Top. **1** (2008), 527-550.

[37] —, Decorated Teichmüller Theory, European Mathematical Society, Zürich (2012).

[38] M. Pillsbury, H. Orland, A. Zee, *Steepest descent calculation of RNA pseudoknots*, Phys. Rev. E **72** (2005).

[39] M. Pillsbury, J. A. Taylor, H. Orland, A. Zee, *An Algorithm for RNA Pseudoknots*, (2005), arXiv: cond-mat/0310505v2,

[40] C. M. Reidys, Combinatorial Computational Biology of RNA, Springer, New York (2011).

[41] C. M. Reidys, F. W. D. Huang, J. E. Andersen, R. C. Penner, P. F. Stadler, M. E. Nebel, *Topology and prediction of RNA pseudoknots*, Bioinformatics **27** (2011), 1076-1085.

[42] G. Vernizzi, H. Orland, A. Zee, *Enumeration of RNA structures by matrix models*, Phys. Rev. Lett. **94**, 16.

[43] Wolfram Research, Inc., Mathematica, Version 8.0, Champaign, IL (2010).

[44] S. Zhu, *On a proof of the Bouchard-Sulkowski conjecture*, 1108.2831 [math.AG].