ENUMERATION OF CHORD DIAGRAMS ON MANY INTERVALS AND THEIR NON-ORIENTABLE ANALOGS

N. V. ALEXEEV, J. E. ANDERSEN, R. C. PENNER, AND P. G. ZOGRAF

Abstract. Two types of connected chord diagrams with chord endpoints lying in a collection of ordered and oriented real segments are considered here: the real segments may contain additional bivalent vertices in one model but not in the other. In the former case, we record in a generating function the number of fatgraph boundary cycles containing a fixed number of bivalent vertices while in the latter, we instead record the number of boundary cycles of each fixed length. Second order, non-linear, algebraic partial differential equations are derived which are satisfied by these generating functions in each case giving efficient enumerative schemes. Moreover, these generating functions provide multi-parameter families of solutions to the KP hierarchy. For each model, there is furthermore a non-orientable analog, and each such model likewise has its own associated differential equation. The enumerative problems we solve are interpreted in terms of certain polygon gluings. As specific applications, we discuss models of several interacting RNA molecules. We also study a matrix integral which computes numbers of chord diagrams in both orientable and non-orientable cases in the model with bivalent vertices, and the large-N limit is computed using techniques of free probability.

1. Introduction

A *partial chord diagram* is a connected fatgraph (i.e., a graph equipped with a cyclic order on the half edges incident to each vertex) comprised of an ordered set of $b \geq 1$ disjoint real line segments (called *backbones*) connected with $k \geq 0$ *chords* in the upper half plane with distinct endpoints, so that there are $2k$ vertices of degree three (or *chord endpoints*) and $l \geq 0$ vertices of degree two (or *marked points*) all belonging to the backbones (in effect, ignoring the vertices of degree one arising from backbone endpoints.) If $l = 0$ so there are no marked points, then we call the diagram a *(complete) chord diagram*. Each partial or complete chord diagram is a spine of an orientable surface with $n \geq 1$ boundary components and therefore has a well-defined topological genus. The genus g of a partial chord diagram on b backbones and its number n of boundary components are related by Euler's formula $b - k + n = 2 - 2g$.

Chord diagrams occur pervasively in mathematics, which further highlights the importance of the counting results obtained here. To mention a few, see the theory of finite type invariants of knots and links [19] (cf. also [9]), the representation theory of Lie algebras [12], the geometry of moduli spaces of flat connections on

Research supported by the Centre for Quantum Geometry of Moduli Spaces which is funded by the Danish National Research Foundation. The work of NA and PZ was further supported by the Government of the Russian Federation megagrant 11.G34.31.0026, by JSC "Gazprom Neft", and by the RFBR grants 13-01-00935-a and 13-01-12422-OFI-M2. In addition, NA was partially supported by the RFBR grant 14-01-00500-a and by the SPbSU grant 6.38.672.2013, and PZ was partially supported by the RFBR grant 14-01-00373-a.

surfaces [5, 6] and mapping class groups [2]. Moreover and as we shall further explain later, partial and complete chord diagrams each provide a useful model [28, 29, 25, 37] for the combinatorics of interacting RNA molecules with the associated genus filtration of utility in enumerative problems [3, 11, 25, 32, 33, 37, 38] and in folding algorithms on one [35, 10] and two backbones [4].

Our goal is to enumerate various classes of connected partial and complete chord diagrams, and to this end, we next introduce combinatorial parameters, where each enumerative problem turns out to be solved by an elegant partial differential equation on a suitable generating function in dual variables. In effect, creation and annihilation operators for the combinatorial data are given by multiplication and differentiation operators in the dual variables leading to algebraic differential equations.

We say that a partial chord diagram has

- *backbone spectrum* $\boldsymbol{b} = (b_1, b_2, \ldots)$ if the diagram has b_i backbones with precisely $i \geq 1$ vertices (of degree either two or three);
- *boundary point spectrum* $\boldsymbol{n} = (n_0, n_1, \ldots)$ if its boundary contains n_i connected components with i marked points;
- *boundary length spectrum* $\boldsymbol{p} = (p_1, p_2, \ldots)$ if the boundary cycles of the diagram consist of p_i edge-paths of length $i \geq 1$, where the *length* of a boundary cycle is the number of chords it traverses counted with multiplicity (as usual on the graph obtained from the diagram by collapsing each backbone to a distinct point) *plus* the number of backbone undersides it traverses (or in other words, the number of traversed backbone intervals obtained by removing all the chord endpoints from all the backbones).

The data $\{g, k, l; \boldsymbol{b}; \boldsymbol{n}; \boldsymbol{p}\}$ is called the *type* of a partial chord diagram (cf. Fig. 1). Note that the entries in the data set $\{g, k, l; \boldsymbol{b}; \boldsymbol{n}; \boldsymbol{p}\}$ are not independent. In particular, we have

$$b = \sum_{i\geq 1} b_i, \quad n = \sum_{i\geq 0} n_i = \sum_{i\geq 1} p_i, \quad l = \sum_{i>0} i n_i,$$
$$2k + l = \sum_{i>0} i b_i, \quad 2k + b = \sum_{i\geq 1} i p_i.$$

Let $\mathcal{N}_{g,k,l}(\boldsymbol{b}, \boldsymbol{n}, \boldsymbol{p})$ denote the number of distinct connected partial chord diagrams of type $\{g, k, l; \boldsymbol{b}; \boldsymbol{n}; \boldsymbol{p}\}$ taken to be zero if there are no chord diagrams of the specified type. Our two basic models involve boundary point spectra of partial chord diagrams and boundary length spectra of complete chord diagrams, and each basic model, in turn, has both an orientable and a non-orientable incarnation.

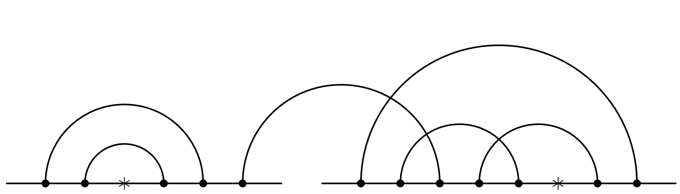

FIGURE 1. The partial chord diagram of type $\{1, 6, 2; \boldsymbol{e}_6 + \boldsymbol{e}_8; 2\boldsymbol{e}_0 + 2\boldsymbol{e}_1; \boldsymbol{e}_1 + 2\boldsymbol{e}_2 + \boldsymbol{e}_9\}$. Here $\boldsymbol{e}_i$ stands for the sequence with 1 in the i-th place and 0 elsewhere.

We may also consider non-orientable chord diagrams. Let $\widetilde{\mathcal{N}}_{h,k,l}(\boldsymbol{b},\boldsymbol{n},\boldsymbol{p})$ denote the number of both orientable and non-orientable connected diagrams on b backbones, out of which exactly b_i have i vertices, with k pairs of vertices connected by (twisted or untwisted) chords, with boundary point and boundary length spectra $\boldsymbol{n}$ and $\boldsymbol{p}$ respectively, and with Euler characteristic $2-h-n,\ n=\sum_{i=1}^{\infty} n_i$, where h denotes twice the genus in the orientable case and the number of cross caps in the non-orientable case. This can evidently be formalized in the language of planar projections of chord diagrams by two-coloring the chords depending upon whether they preserve or reverse the orientation of the plane of projection.

For partial chord diagrams and boundary point spectra, we shall count the subsets

$$\mathcal{N}_{g,k,l}(\boldsymbol{b},\boldsymbol{n})=\sum_{\boldsymbol{p}} \mathcal{N}_{g,k,l}(\boldsymbol{b},\boldsymbol{n},\boldsymbol{p})$$

in the orientable case and

$$\widetilde{\mathcal{N}}_{h,k,l}(\boldsymbol{b},\boldsymbol{n})=\sum_{\boldsymbol{p}} \widetilde{\mathcal{N}}_{h,k,l}(\boldsymbol{b},\boldsymbol{n},\boldsymbol{p})$$

in the non-orientable case.

We can equivalently replace each backbone component containing b_i vertices by a polygon with b_i sides (one of which is distinguished, corresponding to the first along the backbone). Thus, the numbers $\mathcal{N}_{g,k,l}(\boldsymbol{b},\boldsymbol{n})$ count the orientable genus $g=1+(k-b-n)/2$ connected gluings of b polygons, among which exactly b_i have i sides, with k pairs of sides identified in such a way that the boundary of the glued surface has exactly n_i connected components consisting of i sides.[1] In particular, the Harer-Zagier numbers $\epsilon_{g,k}$ [15] that enumerate closed orientable genus g gluings of a $2k$-gon coincide with the numbers $\mathcal{N}_{g,k,l}(\boldsymbol{n},\boldsymbol{b})$ with $\boldsymbol{n}=(k-2g+1,0,0,\dots)$ and $\boldsymbol{b}=\boldsymbol{e}_{2k}$, where we denote by $\boldsymbol{e}_i$ the vector with 1 in the i-th place and 0 elsewhere.

A useful notation for exponentiating a m-tuple $\boldsymbol{a}=(a_1,a_2,\dots,a_m)$ of variables by an integral m-tuple $\boldsymbol{\alpha}=(\alpha_1,\alpha_2,\dots,\alpha_m)$ is to write simply $\boldsymbol{a}^{\boldsymbol{\alpha}}=a_1^{\alpha_1}a_2^{\alpha_2}\dots a_m^{\alpha_m}$; we extend this notation in case $\boldsymbol{a}=(a_1,a_2,a_3,\dots)$ is a fixed infinite sequence of variables and $\boldsymbol{\alpha}$ is a finite tuple. In this notation and setting $\boldsymbol{s}=(s_0,s_1,\dots)$ and $\boldsymbol{t}=(t_1,t_2,\dots)$, we define the orientable, multi-backbone, boundary point spectrum generating function $F(x,y;\boldsymbol{s};\boldsymbol{t})=\sum_{b\ge 1} F_b(x,y;\boldsymbol{s};\boldsymbol{t})$, where

$$F_b(x,y;\boldsymbol{s};\boldsymbol{t})=\frac{1}{b!}\sum_{k=b-1}^{\infty}\sum_{\boldsymbol{n}}\sum_{\sum b_i=b}\mathcal{N}_{g,k,l}(\boldsymbol{b},\boldsymbol{n})x^{2g-2}y^k\boldsymbol{s}^{\boldsymbol{n}}\boldsymbol{t}^{\boldsymbol{b}}\ , \tag{1}$$

and the non-orientable generating functions $\tilde{F}(x,y;\boldsymbol{s};\boldsymbol{t})=\sum_{b\ge 1}\tilde{F}_b(x,y;\boldsymbol{s};\boldsymbol{t})$, where

$$\tilde{F}_b(x,y;\boldsymbol{s};\boldsymbol{t})=\frac{1}{b!}\sum_{k=b-1}^{\infty}\sum_{\boldsymbol{n}}\sum_{\sum b_i=b}\widetilde{\mathcal{N}}_{g,k,l}(\boldsymbol{b},\boldsymbol{n})x^{h-2}y^k\boldsymbol{s}^{\boldsymbol{n}}\boldsymbol{t}^{\boldsymbol{b}} \tag{2}$$

(we recall that $2-2g=b-k+\sum_{i\ge 0} n_i$ in the orientable case and $2-h=b-k+\sum_{i\ge 0} n_i$ in the non-orientable case, while in the both cases $l=\sum_{i\ge 1} in_i$).

[1]For $b=1$, a similar problem was addressed in [1], but the formulas derived there are considerably different from the ones obtained in this paper and do not stand a numerical check.

Theorem 1 (Boundary point spectrum for partial chord diagrams)**.** *Consider the linear differential operators*

$$L_0 = \frac{1}{2}\sum_{i=0}^{\infty}\sum_{j=0}^{i}(i+2)s_j s_{i-j}\frac{\partial}{\partial s_{i+2}}\,,$$

$$L_1 = \frac{1}{2}\sum_{i=0}^{\infty}(i+2)(i+1)s_i\frac{\partial}{\partial s_{i+2}}\,,$$

$$L_2 = \frac{1}{2}\sum_{i=2}^{\infty}s_{i-2}\sum_{j=1}^{i-1}j(i-j)\,\frac{\partial^2}{\partial s_j\partial s_{i-j}}$$

and the quadratic differential operator

$$QF = \frac{1}{2}\sum_{i=2}^{\infty}s_{i-2}\sum_{j=1}^{i-1}j(i-j)\,\frac{\partial F}{\partial s_j}\cdot\frac{\partial F}{\partial s_{i-j}}\,.$$

Then the following partial differential equations hold:

$$\frac{\partial F_1}{\partial y} = (L_0 + x^2 L_2)F_1\ , \qquad \frac{\partial \tilde{F}_1}{\partial y} = (L_0 + xL_1 + 2x^2 L_2)\tilde{F}_1\,,$$

$$\frac{\partial F}{\partial y} = (L_0 + x^2 L_2 + x^2 Q)F\ , \qquad \frac{\partial \tilde{F}}{\partial y} = (L_0 + xL_1 + 2x^2 L_2 + 2x^2 Q)\tilde{F}\,.$$

These equations, together with the common for each case initial condition at $y = 0$ given by $x^{-2}\sum_{i\geq 1}s_i t_i$, determine the generating functions $F_1, F, \tilde{F}_1, \tilde{F}$ uniquely.

Equivalently, each differential equation is solved by exponentiating k times the operator on the right hand side applied to $x^{-2}\sum_{i\geq 1}s_i t_i$, for example,

$$F_1(x, y; s_0, s_1, \ldots; t_1, t_2, \ldots) = e^{y(L_0 + x^2 L_2)}\left(x^{-2}\textstyle\sum_{i\geq 1}s_i t_i\right)\,,$$

$$e^{F(x,y;s_0,s_1,\ldots;t_1,t_2,\ldots)} = e^{y(L_0+x^2L_2)}e^{x^{-2}\sum_{i\geq 1}s_i t_i}\,.$$

This explains the relationship between these differential equations and the corresponding enumerative problems. These are the most efficient enumerations of which we are aware. As we shall see in the proof, each term corresponds to adding a certain type of chord: L_0 and L_2, respectively, for chords with both endpoints on the same and different boundary components lying in a common component, Q for chords whose removal separates the diagram, and L_1 the analogue of L_0 for Möbius bands that give rise to Möbius graphs as compared to fatgraphs in the oriented case (the subscripts 0,1 and 2 by L reflect the change in the Euler characteristic of the chord diagram under such an operation).

In the last section of this paper, we provide matrix model formulas for certain linear combinations of the numbers $\mathcal{N}_{g,k,l}(\boldsymbol{b},\boldsymbol{n})$ and $\widetilde{\mathcal{N}}_{h,k,l}(\boldsymbol{b},\boldsymbol{n})$. This allows us to compare our computations for partial chord diagrams with results on a certain limiting spectral distribution, the so-called large N-limit for one backbone. Note that a recursion for the numbers $\widetilde{\mathcal{N}}_{h,k,0}(\boldsymbol{e}_{2k},\boldsymbol{n})$, for $\boldsymbol{n} = (k-h+1, 0, \ldots)$, of all complete (not necessarily orientable) gluings of a $2k$-gon was derived in [21] using the methods of random matrix theory. Our formulas specialize to those of [21] in this particular case.

For complete chord diagrams and boundary length spectra, we shall count the subsets

$$\mathcal{N}_{g,k,b}(\boldsymbol{p}) = \sum_{\sum b_i = b} \sum_{\boldsymbol{n}} \mathcal{N}_{g,k,0}(\boldsymbol{b},\boldsymbol{n},\boldsymbol{p})$$

in the orientable case and

$$\widetilde{\mathcal{N}}_{g,k,b}(\boldsymbol{p}) = \sum_{\sum b_i = b} \sum_{\boldsymbol{n}} \widetilde{\mathcal{N}}_{h,k,0}(\boldsymbol{b},\boldsymbol{n},\boldsymbol{p})$$

in the non-orientable case. We define the orientable, multi-backbone, length spectrum generating function $G(x,y,t;\boldsymbol{s}) = \sum_{b\geq 1} G_b(x,y;\boldsymbol{s})t^b$, where

$$G_b(x,y;\boldsymbol{s}) = \frac{1}{b!}\sum_{k=0}^{\infty}\sum_{\boldsymbol{p}} \mathcal{N}_{g,k,b}(\boldsymbol{p})x^{2g-2}y^k\boldsymbol{s}^{\boldsymbol{p}}, \tag{3}$$

and the non-orientable generating function $\tilde{G}(x,y,t;\boldsymbol{s}) = \sum_{b\geq 1} \tilde{G}_b(x,y;\boldsymbol{s})t^b$, where

$$\tilde{G}_b(x,y;\boldsymbol{s}) = \frac{1}{b!}\sum_{k\geq 0}\sum_{\boldsymbol{p}} \widetilde{\mathcal{N}}_{h,k,b}(\boldsymbol{p})x^{h-2}y^k\boldsymbol{s}^{\boldsymbol{p}}\,. \tag{4}$$

Theorem 2 (Boundary length spectrum for complete chord diagrams)**.** *Define the linear differential operators*

$$\begin{aligned} K_0 &= \frac{1}{2}\sum_{i=3}^{\infty}\sum_{j=1}^{i-1}\,(i-2)s_j s_{i-j}\frac{\partial}{\partial s_{i-2}}\ , \\ K_1 &= \frac{1}{2}\sum_{i=3}^{\infty}\,(i-2)(i-1)\,s_i\frac{\partial}{\partial s_{i-2}}\,, \\ K_2 &= \frac{1}{2}\sum_{i=2}^{\infty}\sum_{j=1}^{i-1}\,j(i-j)\ s_{i+2}\frac{\partial^2}{\partial s_j\partial s_{i-j}} \end{aligned} \tag{5}$$

and the quadratic differential operator

$$RG = \frac{1}{2}\sum_{i=2}^{\infty}s_{i+2}\sum_{j=1}^{i-1}\,j(i-j)\,\frac{\partial G}{\partial s_j}\cdot\frac{\partial G}{\partial s_{i-j}}\ . \tag{6}$$

Then the following partial differential equations hold:

$$\frac{\partial G_1}{\partial y} = (K_0 + x^2K_2)G_1\ , \qquad \frac{\partial \tilde{G}_1}{\partial y} = (K_0 + xK_1 + 2x^2K_2)\tilde{G}_1\,,$$
$$\frac{\partial G}{\partial y} = (K_0 + x^2K_2 + x^2R)G\ , \qquad \frac{\partial \tilde{G}}{\partial y} = (K_0 + xK_1 + 2x^2K_2 + 2x^2R)\tilde{G}\,.$$

These equations, together with the common in each case initial condition at $y=0$ given by $x^{-2}ts_1$, determine the generating functions $G_1, G, \tilde{G}_1, \tilde{G}$ uniquely.

Remark 1. Complete gluing of a $2k$-gon with a marked edge can be enumerated in a similar way. Consider the image of the polygon perimeter, that is, the graph embedded in the glued surface. We say that the embedded graph has the *vertex spectrum* $\boldsymbol{v} = (v_1, v_2, \dots)$ if there are exactly v_i vertices of degree i. Let $\widehat{N}_{g,k}(\boldsymbol{v})$

denote the number of genus g orientable gluings of a $2k$-gon, such that the embedded graph has the vertex spectrum $\boldsymbol{v}$. The generating function

$$\widehat{F}(x,y;s_1,s_2,\dots)=\sum_{k=1}^{\infty}\sum_{g=0}^{[k/2]}\sum_{\boldsymbol{v}}\widehat{N}_{g,k}(\boldsymbol{v})\,x^{2g-2}y^{k-1}\boldsymbol{s}^{\boldsymbol{v}} \tag{7}$$

for the numbers $\widehat{N}_{g,k}(\boldsymbol{v})$ satisfies the equation

$$\frac{\partial\widehat{F}(x,y;\boldsymbol{s})}{\partial y}=(K_0+x^2K_2)\widehat{F}(x,y;\boldsymbol{s}), \tag{8}$$

and is uniquely determined by it together with the initial condition

$$\widehat{F}\,|_{y=0}\;=x^{-2}s_1^2.$$

Actually, $\widehat{F}$ and G_1 are explicitly related by the formula

$$\frac{1}{2}\Lambda_1\widehat{F}=\frac{\partial G_1}{\partial y}\,,\qquad \Lambda_1=\sum_{i=1}^{\infty}is_{i+1}\frac{\partial}{\partial s_i} \tag{9}$$

(this immediately follows from the fact that both K_0 and K_2 commute with Λ_1). The same problem, but differently formulated (namely, the enumeration of genus g fatgraphs on n vertices of specified degrees) was recently solved in [13]. However, our generating function (7) for these numbers and the partial differential equation (8) it satisfies are different from their counterparts in [13].

The following observation we learned from M. Kazarian [17, 18]: for $x=1$ the generating functions $F\,|_{x=1}$ and $G\,|_{x=1}$ satisfy an infinite system of non-linear partial differential equations called the KP (Kadomtsev-Petviashvili) hierarchy (in particular, this means that the numbers $\mathcal{N}_{g,k,l}(\boldsymbol{b},\boldsymbol{n})$ and $\mathcal{N}_{g,k,b}(\boldsymbol{p})$ additionally obey an infinite system of recursions). The KP hierarchy is one of the best studied completely integrable systems in mathematical physics. Below are the several first equations of the hierarchy:

$$\begin{aligned}
F_{22}&=-\frac{1}{2}\,F_{11}^2+F_{31}-\frac{1}{12}\,F_{1111}\;,\\
F_{32}&=-F_{11}F_{21}+F_{41}-\frac{1}{6}F_{2111}\;,\\
F_{42}&=-\frac{1}{2}\,F_{21}^2-F_{11}F_{31}+F_{51}+\frac{1}{8}\,F_{111}^2+\frac{1}{12}\,F_{11}F_{1111}-\frac{1}{4}\,F_{3111}+\frac{1}{120}\,F_{111111}\;,\\
F_{33}&=\frac{1}{3}\,F_{11}^3-F_{21}^2-F_{11}F_{31}+F_{51}+\frac{1}{4}\,F_{111}^2+\frac{1}{3}\,F_{11}F_{1111}-\frac{1}{3}\,F_{3111}+\frac{1}{45}\,F_{111111}\;,
\end{aligned} \tag{10}$$

where the subscript i stands for the partial derivative with respect to s_i. The exponential e^F of any solution is called a *tau function* of the hierarchy. The space of solutions (or the space of tau functions) has a nice geometric interpretation as an infinite-dimensional Grassmannian (called the *Sato Grassmannian*), see, e. g., [22] or [17] for details. See also [8] for another application of the Sato Grassmannian to conformal field theory. The space of solutions is homogeneous: there is a Lie algebra $\widehat{\mathfrak{gl}(\infty)}$ (a central extension of $\mathfrak{gl}(\infty)$) that acts infinitesimally on the space of solutions, and the action of the corresponding Lie group is transitive.

Introduce the standard bosonic creation-annihilation operators

$$a_i = \begin{cases} s_i & \text{if } i > 0 \\ 0 & \text{if } i = 0 \\ (-i)\frac{\partial}{\partial s_{-i}} & \text{if } i < 0 \end{cases}$$

and put

$$\Lambda_m = \frac{1}{2}\sum_{i=-\infty}^{\infty} a_i\, a_{m-i}\,,$$
$$M_m = \frac{1}{6}\sum_{i,j=-\infty}^{\infty} :a_i\, a_j\, a_{m-i-j}:$$

(the notation $:a_{i_1}\dots a_{i_r}:$ stands for the ordered product $a_{i_{\sigma(1)}}\dots a_{i_{\sigma(r)}}$, where $\sigma \in S_r$ is a permutation such that $i_{\sigma(1)} \geq \dots \geq i_{\sigma(r)}$).[2] All the operators $a_i,\ \Lambda_m,\ M_m$ belong to the Lie algebra $\widehat{\mathfrak{gl}(\infty)}$. Moreover, it is easy to check that

$$L_0 + L_2 = s_0^2\frac{\partial}{\partial s_2} + s_0\Lambda_{-2} + M_{-2}\,, \qquad K_0 + K_2 = M_2\,, \tag{11}$$

so that $L_0 + L_2$ and $K_0 + K_2$ also belong to $\widehat{\mathfrak{gl}(\infty)}$. Now we notice that the exponentials $e^{\sum s_i t_i}$ and e^{ts_1} of the initial conditions in Theorems 1 and 2 both are KP tau functions for a trivial reason – their logarithms are linear in $s_1, s_2, \dots$ and therefore obviously satisfy the equations of KP hierarchy (10) for any values of the other parameters. Moreover, both $e^{L_0+L_2}$ and $e^{K_0+K_2}$ preserve the Sato Grassmannian and map KP tau functions to KP tau functions. Thus, $e^{F|_{x=1}} = e^{L_0+L_2}e^{\sum_{i\geq 1} s_i t_i}$ and $e^{G|_{x=1}} = e^{K_0+K_2}e^{ts_1}$ are KP tau function as well, and we get

Corollary 1 (M. Kazarian [18])**.** *The generating functions*

$$F\,|_{x=1} = F(1, y; s_0, s_1, \dots; t_1, t_2, \dots) \text{ and } G\,|_{x=1} = G(1, y, t; s_1, s_2, \dots)$$

satisfy the infinite system of KP equations (10) with respect to $s_1, s_2, \dots$ for any values of the parameters $y, t, t_1, t_2, \dots$. Equivalently, the partition functions $e^{F|_{x=1}}$ and $e^{G|_{x=1}}$ are (multi-parameter) families of KP tau functions.

Let us now comment on the relevance of the above results to describing the RNA interactions. Define $C_{g,b,k} = \sum_{\boldsymbol{p}} \mathcal{N}_{g,k,b}(\boldsymbol{p})$ to be the number of complete and connected chord diagrams of genus g on b ordered and oriented backbones with k chords, so in particular, $C_{g,1,k}$ is the Harer-Zagier number $\epsilon_{g,k}$. These chord diagrams provide the basic model for a complex of interacting RNA molecules, one RNA molecule for each backbone and one chord for each Watson-Crick[3] bond between nucleic acids, where one demands that the chord endpoints respect the natural ordering[4] of the nucleic acids in each molecule, i.e, in each oriented backbone. It is very natural, as is the attention to connected chord diagrams in order to

[2] The operator M_0 is the famous cut-and-join operator [14] used in the computation of Hurwitz numbers.

[3] These are the allowed bonds G-C and A-U between nucleic acids. For the expert, let us emphasize that any other model including wobble G-U or further exotic base pairs is handled in exactly the same way with one chord for each allowed type of bond.

[4] From the so-called 5' to 3' end as determined by the chemical structure of the RNA.

avoid separate molecular interactions. In reality, RNA folds according to a *partial* chord diagram, i.e., there are in practice unbonded nucleic acids.[5]

Recall from [7] that a shape is a special connected and complete chord diagram which has no parallel chords, has a unique "rainbow" on each backbone, i.e., a chord whose endpoints are closer to the backbone endpoints than any other chord and no "1-chords" connecting vertices consecutive in a single backbone unless the 1-chord is a rainbow. In the very special (genus zero on one backbone) case, the single-chord diagram is permitted since the 1-chord is a rainbow, but in all other cases, there are no 1-chords, each backbone has a unique rainbow, and $p_1 = 0$, $p_2 = b$. If a shape is not the special single-chord diagram and we remove its b rainbows, then the resulting diagram has $p_1 = 0 = p_2$. Conversely, in a chord diagram with $p_1 = 0 = p_2$, no backbone has a rainbow, and b rainbows can be added to produce a shape. Let $S_{g,b,k}$ denote the number of shapes of genus g on b backbones with k chords.

Define the generating functions $C(x, y, t) = \sum_{g=0}^{\infty} \sum_{b=1}^{\infty} C_{g,b}(y) x^{2g-2} t^b$, with

$$C_{g,b}(y) = \frac{1}{b!} \sum_{k=2g+b-1}^{\infty} C_{g,b,k}\, y^k,$$

and $S(x, y, t) = \sum_{g=0}^{\infty} \sum_{b=1}^{\infty} S_{g,b}(y) x^{2g-2} t^b$, with

$$S_{g,b}(y) = \frac{1}{b!} \sum_{k=2g+2b-1}^{6g-6+5b} S_{g,b,k} y^k.$$

It follows by construction that

$$C(x, y, t) = F(x, y; 1, 0, 0, \ldots; t, t, \ldots) = G(x, y, t; 1, 1, \cdots)$$

and $S(x, y, t) = 1 + G(x, y, t; 0, 0, 1, 1, \ldots)$, so we have computed here both the complete chord diagrams $C_{g,b,k}$ and the shapes $S_{g,b,k}$[6]. In fact [3], the generating functions for shapes and chord diagrams are algebraically related by

$$C_{g,b}(z) = \left(\frac{1}{zC_0(z)}\right)^b \; S_{g,b}\left(\frac{C_0(z) - 1}{2 - C_0(z)}\right),$$
$$S_{g,b}(z) = \left(\frac{z}{1 + 2z}\right)^b \; C_{g,b}\left(\frac{z(1 + z)}{(1 + 2z)^2}\right),$$

where $C_0(z) = \frac{1-\sqrt{1-4z}}{2z}$ is the Catalan generating function, the former equation expressing the formal power series $C_{g,b}(z)$ in terms of the polynomial $S_{g,b}(z)$. As a further interesting open problem, inspired by the results of this paper, we ask if there is a non-zero finite order differential operator in the variables (x, y, t) which together with an initial condition determines $C(x, y, t)$?

One point about shapes is that standard combinatorial techniques allow their "inflation" to complete chord diagrams as indicated in the previous formulas, and furthermore, complete chord diagrams can likewise be inflated to partial chord diagrams, cf. [7, 34]. Another point is that shape inflation is well-suited to the

[5]Typically, 50 to 80 percent of nucleic acids participate in Watson-Crick base pairs together with several percent exotic. On the other hand in an extreme example, roughly 50 percent are Watson-Crick and 40 percent exotic for ribosomal RNA.

[6]Furthermore, the free energy $F_g(s, t)$ for the matrix model in [3] is given (up to a constant depending only on g times N^{2-2g}) by our $G(N^{-1}, t^{\frac{1}{2}}, s; 1, 1, 1...)$.

accepted Ansatz for free energy and so provides efficient polynomial-time algorithms for computing minimum free energy RNA folds [35, 34] at least in the planar case. A further geometric point [7] is that shapes of genus g on b backbones are dual to cells in the Harer-Mumford-Strebel [36] or Penner [27] decomposition of Riemann's moduli space of genus g surfaces with b boundary components provided $2g-2+b>0$.

As was already discussed, it is really partial chord diagrams that actually describe complexes of RNA molecules with its distillation first to complete chord diagrams and then to shapes. All three formulations of the combinatorics have thus been treated here, namely, shapes and complete chord diagrams by the previous formulas and partial chord diagrams by inflation or instead directly with our generating function in Theorem 1.

This paper is organized as follows. Section 2 contains basic combinatorial results on the boundary point spectra of chord diagrams on one backbone and derives the equation given before on F_1 (Proposition 2), and section 3 extends these results to include possibly separating edges and derives the equation given before on F (Proposition 4). Boundary point spectra of non-orientable surfaces are discussed in Section 4, and the equations given before on $\tilde{F}_1$ and $\tilde{F}$ are derived (Proposition 5), so together Propositions 2-5 comprise Theorem 1. Section 5 is dedicated to boundary length spectra, and the situation is similar to boundary point spectra in that each counts data for each fatgraph boundary cycle. For this reason, the arguments are only sketched for boundary length spectra culminating in the equations from before on G_1, G, $\tilde{G}_1$ and $\tilde{G}$ (Theorem 2). Section 6 introduces a random matrix technique for partial chord diagrams and provides a matrix integral for boundary point spectra computations in both the orientable and non-orientable cases. Free probability techniques permit the computation of the large-N limit which reproduces computations based on the partial differential equations, providing a consistency check on the entire discussion.

2. Combinatorics of connected partial chord diagrams

As before, $\boldsymbol{e}_i$ denotes the sequence $(0,\dots,0,1,0,\dots)$ with 1 in the i-th place and 0 elsewhere. We say simply that a diagram is of type $\{g,k,l;\boldsymbol{b};\boldsymbol{n}\}$ if it is of type $\{g,k,l;\boldsymbol{b},\boldsymbol{n},\boldsymbol{p}\}$ for some $\boldsymbol{p}$ and let $\mathcal{N}_{g,k,l}(\boldsymbol{b},\boldsymbol{n})=0$ if there are no diagrams of type $\{g,k,l;\boldsymbol{b};\boldsymbol{n}\}$.

Proposition 1. *The numbers $\mathcal{N}_{g,k,l}(\boldsymbol{e}_{2k+l},\boldsymbol{n})$ enumerating one backbone chord diagrams of type $\{g,k,l;\boldsymbol{e}_{2k+l};\boldsymbol{n}\}$ obey the following recursion relation:*

$$\begin{aligned}
k\mathcal{N}_{g,k,l}(\boldsymbol{e}_{2k+l},\boldsymbol{n}) = \\
&\frac{1}{2}\sum_{i=0}^{\infty}\sum_{j=0}^{i}(i+2)(n_{i+2}+1)\mathcal{N}_{g,k-1,l+2}(\boldsymbol{e}_{2k+l},\boldsymbol{n}-\boldsymbol{e}_j-\boldsymbol{e}_{i-j}+\boldsymbol{e}_{i+2})+ \\
&\frac{1}{2}\sum_{i=0}^{\infty}\sum_{j=1}^{i+1}j(i+2-j)(n_j+1+\delta_{j,i+2-j}-\delta_{i,j})(n_{i+2-j}+1-\delta_{j,2})\times \\
&\qquad \mathcal{N}_{g-1,k-1,l+2}(\boldsymbol{e}_{2k+l},\boldsymbol{n}+\boldsymbol{e}_j+\boldsymbol{e}_{i+2-j}-\boldsymbol{e}_i).
\end{aligned} \tag{12}$$

Proof. Let us start with a chord diagram of type $\{g, k, l; \boldsymbol{e}_{2k+l}; \boldsymbol{n}\}$. Note that erasing a chord in a diagram, we keep its endpoints as marked points. This yields two possibilities.

The first possibility is that the chord belongs to two distinct boundary components, say, one with j and the other with $i-j$ marked points. After erasing the chord, these two boundary components join into one component with $i+2$ marked points, and the genus of the diagram does not change (see Fig. 2). Thus, one gets a diagram of genus g with $k-1$ chords, $l+2$ marked points and boundary point spectrum $\boldsymbol{n} - \boldsymbol{e}_j - \boldsymbol{e}_{i-j} + \boldsymbol{e}_{i+2}$.

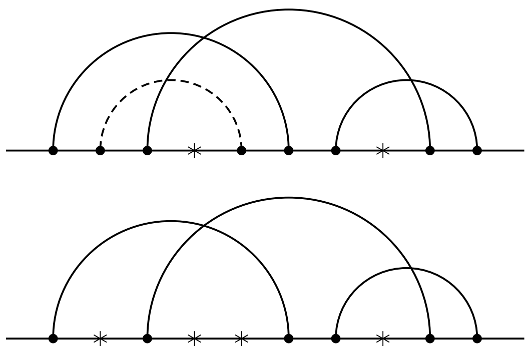

FIGURE 2. Erasing the dashed chord changes the diagram type from $\{1, 4, 2; \boldsymbol{e}_{10}; 2\boldsymbol{e}_0 + \boldsymbol{e}_2\}$ to $\{1, 3, 4; \boldsymbol{e}_{10}; \boldsymbol{e}_0 + \boldsymbol{e}_4\}$

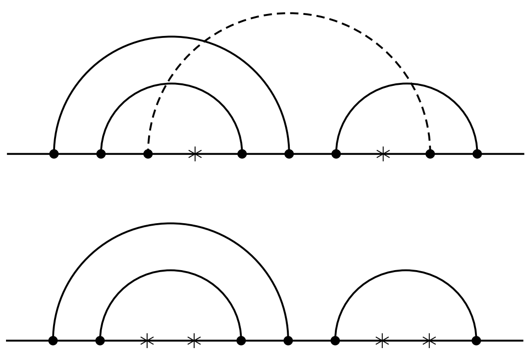

FIGURE 3. Erasing the dashed chord changes the diagram type from $\{1, 4, 2; \boldsymbol{e}_{10}; 2\boldsymbol{e}_0 + \boldsymbol{e}_2\}$ to $\{0, 3, 4; \boldsymbol{e}_{10}; 2\boldsymbol{e}_0 + 2\boldsymbol{e}_2\}$

The second possibility is that one boundary component traverses the chord twice, i.e., once in each direction. Erasing this chord splits the boundary component (say, with i marked points) into two (with j and $i-j+2$ marked points respectively, $0 \le j \le i+1$) (see Fig. 3). In this case, one gets a chord diagram of genus $g-1$ with $k-1$ chords, $l+2$ marked points and boundary point spectrum $\boldsymbol{n} + \boldsymbol{e}_j + \boldsymbol{e}_{i+2-j} - \boldsymbol{e}_i$.

In order to prove (12), let us compute the number of chord diagrams of type $\{g, k, l; \boldsymbol{e}_{2k+l}; \boldsymbol{n}\}$ with one marked chord in two different ways. On the one hand, there are k possibilities to mark a chord in a diagram with k chords, so the number in question is $k\mathcal{N}_{g,k,l}(\boldsymbol{e}_{2k+l},\boldsymbol{n})$. On the other hand, one can join any two marked points with a marked chord on any diagram with $k-1$ chords. We have described above all types of diagrams with $k-1$ chords that could potentially give a k-chord diagram of the required type after adding a chord.

If one takes a diagram of type $\{g,k-1,l+2;\boldsymbol{e}_{2k+l};\boldsymbol{n}-\boldsymbol{e}_j-\boldsymbol{e}_{i-j}+\boldsymbol{e}_{i+2}\}$ (let us assume that $j<i-j$), then there are $n_{i+2}+1$ possibilities to choose a boundary component with $i+2$ marked points. One then needs to connect two marked points on it with a chord in such a way that it splits into two boundary components with j and $i-j$ marked points respectively. This can be done in $(i+2)$ different ways. If $j=i-j$, then there are $\frac{i+2}{2}=j+1$ ways to split the boundary component into two components with j marked points each. For $j>i-j$ we get the same diagrams as in the case $j<i-j$, hence we get the first term on the r.h.s. of (12).

If one takes a diagram of type $\{g-1,k-1,l+2;\boldsymbol{e}_{2k+l};\boldsymbol{n}+\boldsymbol{e}_j+\boldsymbol{e}_{i+2-j}-\boldsymbol{e}_i\}$ (let us assume that $j<i-j+2$), then there are (n_j+1) ways to choose a boundary component with j marked points, provided $j\neq i$. If $j=i$, then $j<2=i+2-j$ and so $j=1=i$ and the number of ways is then n_1. There are $(n_{i-j+2}+1)$ ways to choose a boundary component with $i-j+2$ marked points if $i\neq i+2-j$. If $i+2-j=i$, then the number of ways is n_i. One then needs to connect with a chord a marked point on one boundary component with a marked point on the other one. This can be done in $j(i-j+2)$ different ways. If $j=i-j+2$, then there are $(n_j+2)(n_j+1)/2$ ways to choose a pair of boundary components with j marked points, provided $i\neq j$. If we have $i=j$ and also $j=i-j+2$, then $i=2=j$ and the number of ways is $(n_2+1)n_2/2$. In both cases, there are j^2 ways to connect with a chord two points on different components. This gives us the second term on the r.h.s. of (12). □

Proposition 2. *The one backbone generating function $F_1(x,y;s_0,s_1,\dots;t_1,t_2,\dots)$ is uniquely determined by the equation*

$$\frac{\partial F_1}{\partial y}=LF_1\,,\quad L=L_0+x^2L_2\,, \tag{13}$$

together with the initial condition

$$F_1(x,0;s_0,s_1,\dots;t_1,t_2,\dots)=\frac{1}{x^2}\sum_{i=1}^{\infty}s_it_i\,. \tag{14}$$

Equivalently, we have

$$F_1(x,y;s_0,s_1,\dots;t_1,t_2,\dots)=e^{yL}\left(\frac{1}{x^2}\sum_{i=1}^{\infty}s_it_i\right)\,. \tag{15}$$

Proof. It is straightforward to check that the equation $\frac{\partial F_1}{\partial y}=LF_1$ is equivalent to formula (12). Moreover, every chord diagram of type $\{g,k,l;\boldsymbol{e}_{2k+l};\boldsymbol{n}\}$ can be obtained from the unique diagram of type $\{0,0,2k+l;\boldsymbol{e}_{2k+l};\boldsymbol{e}_{2k+l}\}$ by adding k chords to it. On the level of F_1, this amounts to applying the operator L to $x^{-2}s_{2k+l}t_{2k+l}$ precisely k times and taking the coefficient of the monomial $x^{2g-2}t_{2k+l}s_0^{n_0}s_1^{n_1}\dots$ in $L^k(x^{-2}s_{2k+l}t_{2k+l})$ which is equal to $k!\mathcal{N}_{g,k,l}(\boldsymbol{e}_{2k+l},\boldsymbol{n})$ by formula (12). □

Remark 2. Proposition 2 allows us to compute the numbers $\mathcal{N}_{g,k,l}(\boldsymbol{e}_{2k+l},\boldsymbol{n})$ reasonably quickly. For instance, take $g=0$ and put $s_0=1$, $s_i=qy^is^i$, $i\geq 1$, $t_j=$

$1,\ j \geq 1$. The several first coefficients of $x^2 F_1(x, y^2; 1, qys, q(ys)^2, \dots; 1, 1, \dots)$ in y for $x = 0$ and $k = 0, \dots, 8$ are:

$$\begin{aligned}
k = 0: &\quad 1\\
k = 1: &\quad qs\\
k = 2: &\quad qs^2 + 1\\
k = 3: &\quad qs^3 + 3qs\\
k = 4: &\quad qs^4 + (4q + 2q^2)s^2 + 2\\
k = 5: &\quad qs^5 + (5q + 5q^2)s^3 + 10qs\\
k = 6: &\quad qs^6 + (6q + 9q^2)s^4 + (15q + 15q^2)s^2 + 5\\
k = 7: &\quad qs^7 + (7q + 14q^2)s^5 + (21q + 42q^2 + 7q^3)s^3 + 35qs\\
k = 8: &\quad qs^8 + (8q + 20q^2)s^6 + (28q + 84q^2 + 28q^3)s^4 + (56q + 84q^2)s^2 + 14
\end{aligned}$$

In Section 6, we will derive these same polynomials by matrix integration methods.

3. The multibackbone case

Let us proceed with the multibackbone case.

Proposition 3. *The numbers $\mathcal{N}_{g,k,l}(\boldsymbol{b}, \boldsymbol{n})$ obey the following recursion relation:*

$$\begin{aligned}
&k\,\mathcal{N}_{g,k,l}(\boldsymbol{b}, \boldsymbol{n}) =\\
&\quad \frac{1}{2}\sum_{i=0}^{\infty}\sum_{j=0}^{i}(i+2)(n_{i+2}+1)\,\mathcal{N}_{g,k-1,l+2}(\boldsymbol{b}, \boldsymbol{n} - \boldsymbol{e}_j - \boldsymbol{e}_{i-j} + \boldsymbol{e}_{i+2}) +\\
&\quad \frac{1}{2}\sum_{i=0}^{\infty}\sum_{j=1}^{i+1} j(i+2-j)(n_j + 1 + \delta_{j,i+2-j} - \delta_{i,j})(n_{i+2-j} + 1 - \delta_{j,2})\times\\
&\qquad\qquad \mathcal{N}_{g-1,k-1,l+2}(\boldsymbol{b}, \boldsymbol{n} + \boldsymbol{e}_j + \boldsymbol{e}_{i+2-j} - \boldsymbol{e}_i) +\\
&\quad \frac{1}{2}\sum_{i=0}^{\infty}\sum_{j=1}^{i+1}\sum_{g_1+g_2=g}\ \sum_{k_1+k_2=k-1}\ \sum_{\boldsymbol{n}^{(1)}+\boldsymbol{n}^{(2)}=\boldsymbol{n}-\boldsymbol{e}_i}\ \sum_{\boldsymbol{b}^{(1)}+\boldsymbol{b}^{(2)}=\boldsymbol{b}}\\
&\qquad\qquad j(i+2-j)(n_j^{(1)}+1)(n_{i+2-j}^{(2)}+1)\,\frac{b!}{b^{(1)}!b^{(2)}!}\times\\
&\qquad\qquad \mathcal{N}_{g_1,k_1,l_1+j}(\boldsymbol{b}^{(1)}, \boldsymbol{n}^{(1)} + \boldsymbol{e}_j)\,\mathcal{N}_{g_2,k_2,l_2+i+2-j}(\boldsymbol{b}^{(2)}, \boldsymbol{n}^{(2)} + \boldsymbol{e}_{i+2-j}), \qquad (16)
\end{aligned}$$

where

$$b^{(r)} = \sum_{i=1}^{\infty} b_i^{(r)}, \quad l_r = \sum_{i=1}^{\infty} i n_i^{(r)}, \quad \sum_{i=0}^{\infty} n_i^{(r)} = k_r - 2g_r - b^{(r)} + 2\,, \quad r = 1, 2\,.$$

Proof. The multibackbone case is similar to the one backbone case, and the derivation of the first two sums on the r.h.s. of (16) repeats verbatim the proof of (12), cf. Proposition 1. The only difference is that erasing a chord can split the diagram into two connected components (see Fig. 4).

This possibility is encoded in the 6-fold sum on the r.h.s. of (16). There are exactly

$$j(i+2-j)n_j^{(1)}n_{i+2-j}^{(2)}\frac{b!}{b^{(1)}!b^{(2)}!}$$

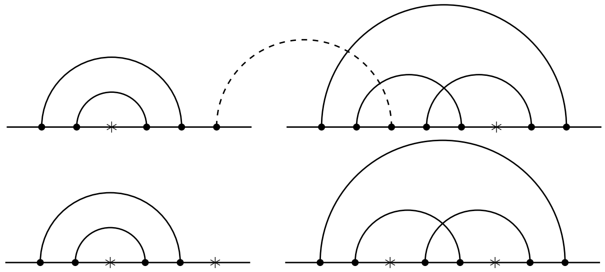

FIGURE 4. Erasing the dashed chord splits the chord diagram of type $\{1,6,2;\boldsymbol{e}_6+\boldsymbol{e}_8;2\boldsymbol{e}_0+2\boldsymbol{e}_1\}$ into two diagrams of types $\{0,2,2;\boldsymbol{e}_6;\boldsymbol{e}_0+2\boldsymbol{e}_1\}$ and $\{1,3,2;\boldsymbol{e}_8;\boldsymbol{e}_0+\boldsymbol{e}_2\}$

ways to get a chord diagram of a type $\{g,k,l;\boldsymbol{b};\boldsymbol{n}\}$ from two diagrams of types $\{g_1,k_1,l_1;\boldsymbol{b}^{(1)};\boldsymbol{n}^{(1)}\}$ and $\{g_2,k_2,l_2;\boldsymbol{b}^{(2)};\boldsymbol{n}^{(2)}\}$. Namely, there are $n_j^{(1)}$ ways to choose a boundary component with j marked points on the first diagram, and there are $n_{i+2-j}^{(2)}$ ways to choose a boundary component with $i-j+2$ marked points on the second diagram. There are $j(i-j+2)$ ways to connect a marked point on the first boundary component with a marked point on the second one. The remaining factor $\frac{b!}{b^{(1)}!b^{(2)}!}$ counts the number of different ordered splittings of a b-backbone diagram into two connected ones that contain $b^{(1)}$ and $b^{(2)}$ backbones respectively. □

Proposition 4. *The generating function $F(x,y;s_0,s_1,\dots;t_1,t_2,\dots)$ is uniquely determined by the equation*

$$\frac{\partial F}{\partial y}=(L+x^2Q)F, \tag{17}$$

where $L=L_0+x^2L_2$ and

$$QF=\frac{1}{2}\sum_{i=2}^{\infty}s_{i-2}\sum_{j=1}^{i-1}j(i-j)\,\frac{\partial F}{\partial s_j}\cdot\frac{\partial F}{\partial s_{i-j}}\ , \tag{18}$$

together with the same initial condition

$$F(x,0;s_0,s_1,\dots;t_1,t_2,\dots)=\frac{1}{x^2}\sum_{i=1}^{\infty}s_it_i\ . \tag{19}$$

Equivalently, the multibackbone partition function e^F satisfies the equation

$$\frac{\partial e^F}{\partial y}=Le^F$$

and is explicitly given by

$$e^{F(x,y;s_0,s_1,\dots;t_1,t_2,\dots)}=e^{yL}\left(e^{\frac{1}{x^2}\sum_{i=1}^{\infty}s_it_i}\right)\ . \tag{20}$$

Proof. As in the one backbone case, a straightforward computation shows that recursion (16) is equivalent to the equation $\frac{\partial F}{\partial y}=(L+x^2Q)F$ (where the 6-fold sum translates into the quadratic term QF). Moreover, every chord diagram of type $\{g,k,l;\boldsymbol{n};\boldsymbol{b}\}$ can be obtained from the disjoint collection of b diagrams of type $\{0,0,i;\boldsymbol{e}_i,\boldsymbol{e}_i\}$ (each taken with multiplicity b_i) by connecting them with k chords.

Let $F^{(k)}(x;s_0,s_1,\dots;t_1,t_2,\dots)$ be the coefficient of y^k in the total generating function F, so

$$F^{(k+1)}(x;s_0,s_1,\dots;t_1,t_2,\dots)$$

is the the coefficient of y^{k+1} in

$$y(L+x^2Q)\left(\sum_{i=0}^{k} y^i F^{(i)}(x;s_0,s_1,\dots;t_1,t_2,\dots)\right),$$

where $F^{(0)}(x;s_0,s_1,\dots;t_1,t_2,\dots)=\frac{1}{x^2}\sum_{i\geq 1}s_it_i$. $\square$

Remark 3. The enumeration problem of *complete* (i.e., giving a closed surface) orientable gluings of two and three polygons (or equivalently chord diagrams on 2 or 3 backbones without marked points) was solved in [7] and independently in [26]by different methods.

4. Non-orientable polygon gluings

This section is dedicated to proving the following result:

Proposition 5. *The one backbone generating function* $\tilde{F}_1(x,y;s_0,s_1,\dots;t_1,t_2,\dots)$ *is uniquely determined by the equation*

$$\frac{\partial \tilde{F}_1}{\partial y}=(L_0+xL_1+2x^2L_2)\tilde{F}_1 \tag{21}$$

together with the initial condition

$$\tilde{F}_1(x,0;s_0,s_1,\dots;t_1,t_2,\dots)=\frac{1}{x^2}\sum_{i=1}^{\infty}s_it_i\ . \tag{22}$$

The generating function $\tilde{F}(x,y;s_0,s_1,\dots;t_1,t_2,\dots)$ *is uniquely determined by the equation*

$$\frac{\partial \tilde{F}}{\partial y}=(L_0+xL_1+2x^2L_2+2x^2Q)\tilde{F}\ , \tag{23}$$

together with the same initial condition (22).

Proof. The non-orientable case is similar to the orientable one. On the combinatorial level, the difference is that when one glues two sides on the same connected component of a boundary with a twist, one adds a cross-cap to the surface without changing the number of boundary components. On the level of the generating function $\tilde{F}_1$, this adds the term $xL_1\tilde{F}_1$ on the r.h.s. of (21). If one glues two sides belonging to distinct components of the boundary, then there is no difference between the twisted and untwisted gluings, so that one just has to count the term $x^2L_2\tilde{F}_1$ twice. The multibackbone generating function is treated analogously. $\square$

Using Proposition 5, we can compute several first numbers $\widetilde{\mathcal{N}}_{h,k,l}(\boldsymbol{b},\boldsymbol{e}_{2k+l})$. Consider, for example, the decagon gluings, i.e., $2k+l=10$. For $x=1$, the coefficients of the generating series $\tilde{F}_1(1, y; \boldsymbol{e}_{10}; s_0, s_1, \dots)$ in y are listed below for $k = 0, 1, \dots, 5$:

$$\begin{aligned}
k = 0: \quad & s_{10}\ , \\
k = 1: \quad & 10s_0s_8 + 10s_1s_7 + 10s_2s_6 + 10s_3s_5 + 5s_4^2 + 45s_8\ , \\
k = 2: \quad & 45s_0^2s_6 + 90s_4s_0s_2 + 90s_3s_1s_2 + 325s_0s_6 + 300s_1s_5 + 285s_2s_4 \\
& + 1050s_6 + 45s_4s_1^2 + 45s_0s_3^2 + 140s_3^2 + 15s_2^3 + 90s_0s_1s_5\ , \\
k = 3: \quad & 1850s_0s_1s_3 + 360s_0^2s_1s_3 + 1000s_0^2s_4 + 360s_0s_1^2s_2 + 900s_0s_2^2 \\
& + 870s_1^2s_2 + 4900s_4s_0 + 4100s_3s_1 + 120s_0^3s_4 + 30s_1^4 \\
& + 180s_0^2s_2^2 + 1920s_2^2 + 8610s_4\ , \\
k = 4: \quad & 1720\, s_0^3s_2 + 2465\, s_0^2s_1^2 + 8890\, s_0^2s_2 + 7940\, s_0s_1^2 + 21930\, s_0s_2 \\
& + 420\, s_0^3s_1^2 + 210\, s_0^4s_2 + 9120\, s_1^2 + 22905\, s_2\ , \\
k = 5: \quad & 42s_0^6 + 386s_0^5 + 2290s_0^4 + 7150s_0^3 + 12143s_0^2 + 8229s_0\ .
\end{aligned}$$

5. Enumeration of chord diagrams with fixed boundary lengths

We will prove Theorem 2 in analogy to Theorem 1 by combinatorial methods. The partial differential equation on G is equivalent to the following

Proposition 6. *The numbers $\mathcal{N}_{g,k,b}(\boldsymbol{p})$ obey the following recursion relation:*

$$\begin{aligned}
& k\,\mathcal{N}_{g,k,b}(\boldsymbol{p}) = \\
& \frac{1}{2}\sum_{i=1}^{\infty}\sum_{j=0}^{i} i(p_i + 1 - \delta_{i,j+1} - \delta_{j,1})\,\mathcal{N}_{g,k-1,b}(\boldsymbol{p} + \boldsymbol{e}_i - \boldsymbol{e}_{j+1} - \boldsymbol{e}_{i-j+1}) + \\
& \frac{1}{2}\sum_{i=2}^{\infty}\sum_{j=1}^{i-1} j(i-j)(p_j + 1)(p_{i-j} + 1 + \delta_{j,i-j})\,\mathcal{N}_{g-1,k-1,b}(\boldsymbol{p} - \boldsymbol{e}_{i+2} + \boldsymbol{e}_j + \boldsymbol{e}_{i-j}) + \\
& \frac{1}{2}\sum_{i=2}^{\infty}\sum_{j=1}^{i-1} j(i-j) \sum_{g_1+g_2=g}\ \sum_{k_1+k_2=k-1}\ \sum_{b_1+b_2=b}\ \sum_{\boldsymbol{p}^{(1)}+\boldsymbol{p}^{(2)}=\boldsymbol{p}-\boldsymbol{e}_{i+2}} \frac{b!}{b_1!b_2!}\times \\
& \qquad (p_j^{(1)} + 1)(p_{i-j}^{(2)} + 1)\,\mathcal{N}_{g_1,k_1,b_1}(\boldsymbol{p}^{(1)} + \boldsymbol{e}_j)\,\mathcal{N}_{g_2,k_2,b_2}(\boldsymbol{p}^{(2)} + \boldsymbol{e}_{i-j})\,. \qquad (24)
\end{aligned}$$

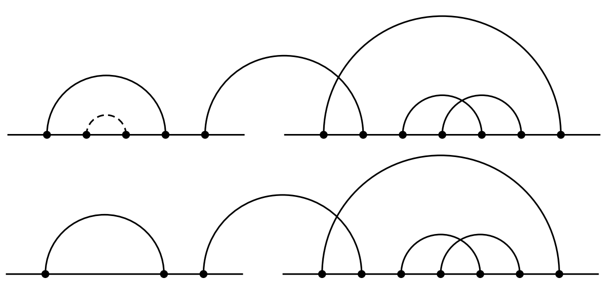

Figure 5. The first backbone has length 6, and the second one has length 8. Erasing the dashed chord joins two boundary components.

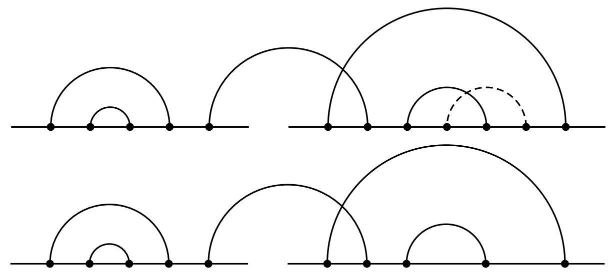

FIGURE 6. Erasing the dashed chord splits a boundary component into two ones.

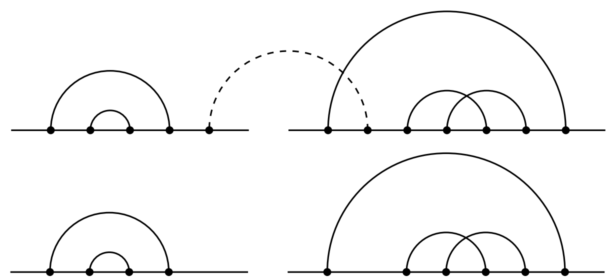

FIGURE 7. Erasing the dashed chord splits the diagram into two connected components.

Proof. The proof is similar to that of Proposition 1. In this case, though, we erase a chord together with its endpoints. There are three possibilities. The first is that the chord belongs to two distinct boundary components (see Fig. 5). Upon erasing the chord, these two components join into one. This possibility is described by the first term on the r.h.s. of (24).

The second possibility occurs when the chord belongs to only one boundary component. When we erase this chord, the boundary component splits into two (see Fig. 6). In this case, the genus of the diagram decreases by 1, and this is described by the second term on the r.h.s. of (24).

In these two cases, the diagram remains connected after erasing a chord. The third possibility occurs when erasing the chord splits the diagram into two connected components (see Fig. 7). This yields the third term on the r.h.s. of (24).

The extension of the proof for the boundary length spectrum in non-orientable case follows a logic similar to that for the boundary point spectrum in Section 4 completing the proof of Theorem 2. □

6. Matrix integral

We show here that certain linear combinations of the numbers $\mathcal{N}_{g,k,l}(\boldsymbol{b},\boldsymbol{n})$ can be interpreted as integrals over the space of Hermitian matrices. Once again, we start with the one backbone case. Let P be a Hermitian $N \times N$ matrix, such that

$P^2 = P$ and $TrP = p$. Consider the matrix integral

$$M_m(s,p,N) = \int_{\mathcal{H}_N} Tr(X+sP)^m \, d\mu(X), \tag{25}$$

where $\mathcal{H}_N$ is the space of Hermitian matrices and

$$d\mu(X) = \frac{1}{vol(\mathcal{H}_N)} \exp(-\frac{1}{2}TrX^2)\, dX$$

is the normalized Gaussian unitary-invariant measure on it (this is a special case of a much more general matrix integral considered in [23]).

Proposition 7. *We have*

$$M_m(s,p,N) = \\ = \sum_{k=0}^{[m/2]} \sum_{g=0}^{[k/2]} \sum_{n_0=0}^{k+1-2g} \sum_{\sum in_i = m-2k} \mathcal{N}_{g,k,m-2k}(\boldsymbol{e}_m, \boldsymbol{n})\, s^{m-2k}\, N^{n_0}\, p^{\sum_{i\geq 1} n_i} \tag{26}$$

Proof. We prove (26) using the Wick formula.

First, note that one can diagonalize the matrix P, and this does not change the measure $d\mu(X)$. Therefore, one can assume that

$$p_{ij} = \begin{cases} 1, & \text{if } i=j, \quad i\leq p, \\ 0, & \text{otherwise.} \end{cases}$$

Second, note that $M_m(s,p,N)$ is a polynomial in s, and the coefficient of s^{m-2k} is $\sum_{\alpha,\beta}\int_{\mathcal{H}_N} Tr(\Pi_{\alpha,\beta})\, d\mu(X)$, where the sum is taken over all products $\Pi_{\alpha,\beta} = X^{\alpha_1}P^{\beta_1}\cdots X^{\alpha_m}P^{\beta_m}$ with $\alpha_i, \beta_i \in \mathbb{Z}_{\geq 0}$ non-negative integers such that $\sum \alpha_i = 2k$ and $\sum \beta_i = m-2k$. We have

$$\int_{\mathcal{H}_N} Tr(\Pi_{\alpha,\beta})\, d\mu(X) = \sum_{i_1=1}^{N}\cdots\sum_{i_m=1}^{N}\int_{\mathcal{H}_N} y_{i_1i_2}y_{i_2i_3}\dots y_{i_mi_1}\, d\mu(X),$$

where

$$y_{i_ji_{j+1}} = \begin{cases} x_{i_ji_{j+1}}, & \text{if } X \text{ is the } j\text{-th factor in the product } \Pi_{\alpha,\beta}, \\ p_{i_ji_{j+1}}, & \text{if } P \text{ is the } j\text{-th factor in the product } \Pi_{\alpha,\beta}. \end{cases}$$

To compute the expectation of the product $y_{i_1i_2}y_{i_2i_3}\dots y_{i_mi_1}$, one has to count all possible matchings between indices of the X-factors. Any product with such a matching can be graphically represented by a chord diagram with k chords and $m-2k$ marked points on the backbone, where the chords correspond to the matched X-factors, and the marked points correspond to P-factors. Each boundary component of the chord diagram is therefore labeled by some index i_j. If there are no marked points on the boundary component, then the corresponding index i_j can take any value from 1 to N. If there are marked points on the boundary component, then the corresponding index i_j can only take values from 1 to p, because p_{ii} is nonzero only when $i \leq p$. Thus, we have

$$\sum_{\alpha,\beta}\int_{\mathcal{H}_N} Tr(\Pi_{\alpha,\beta})\, d\mu(X) = \sum_{g=0}^{[k/2]}\sum_{\sum in_i=m-2k} \mathcal{N}_{g,k,m-2k}(\boldsymbol{e}_m,\boldsymbol{n}) N^{n_0} p^{\sum_{i\geq 1} n_i}$$

which completes the proof. ∎

Let us take the one backbone generating function $F_1(x, y; s_0, s_1, \dots; t_1, t_2, \dots)$ given by (2) and put $x = 1$, $y = 1/z^2$, $s_0 = N$, $s_i = ps^i/z^i$, $i = 1, 2, \dots$, $t_j = 1$, $j = 1, 2, \dots$. This gives us the expectation of the resolvent of $X + sP$:

$$\begin{aligned} &F_1\left(1, \frac{1}{z^2}, N; \frac{ps}{z}, \frac{ps^2}{z^2}, \dots; 1, 1, \dots\right) = \\ &= \sum_{m=0}^{\infty} \sum_{k=0}^{[m/2]} \sum_{g=0}^{[k/2]} \sum_{n_0=0}^{k+1-2g} \sum_{\sum in_i = m-2k} \mathcal{N}_{g,k,m-2k}(\boldsymbol{e}_m, \boldsymbol{n})\, z^{-m}\, N^{n_0}\, p^{\sum_{i\geq 1} n_i} \\ &= -\, z \int_{\mathcal{H}_N} Tr(X + sP - zI)^{-1}\, d\mu(X)\,. \end{aligned} \tag{27}$$

Non-orientable case. The numbers $\widetilde{\mathcal{N}}_{h,k,l}(\boldsymbol{b}, \boldsymbol{n})$ appear as coefficients in the expansion of a matrix integral similarly to the numbers $\mathcal{N}_{g,k,l}(\boldsymbol{b}, \boldsymbol{n})$ again by Wick's Theorem. Namely, consider the matrix integral

$$K_m(s, p, N) = \int_{\mathcal{H}_N(\mathbb{R})} Tr(X + sP)^m\, d\nu(X), \tag{28}$$

where $\mathcal{H}_N(\mathbb{R})$ is the space of real symmetric matrices and

$$d\nu(X) = \frac{1}{vol(\mathcal{H}_N(\mathbb{R}))} \exp(-\frac{1}{2} TrX^2)\, dX$$

is the normalized Gaussian orthogonal-invariant measure on it.

Proposition 8. *We have*

$$\begin{aligned} &K_m(s, p, N) = \\ &= \sum_{k=0}^{[m/2]} \sum_{h=0}^{[k/2]} \sum_{n_0=0}^{k+1-h} \sum_{\sum in_i = m-2k} \widetilde{\mathcal{N}}_{h,k,m-2k}(\boldsymbol{e}_m, \boldsymbol{n})\, s^{m-2k}\, N^{n_0}\, p^{\sum_{i\geq 1} n_i} \end{aligned} \tag{29}$$

Multibackbone case. In the multibackbone case, the matrix integral has a similar form. Take a sequence $\boldsymbol{b} = (b_1, b_2, \dots)$ with a finite number of non-zero elements that are positive integers. Consider the matrix integral

$$I_N(s, p, \boldsymbol{b}) = \int_{\mathcal{H}_N} \prod_{m=0}^{\infty} (Tr(X + sP)^m)^{b_m}\, d\mu(X).$$

This integral is related to the total generating function F by the formula:

$$\frac{1}{b!} I_N(s, p, \boldsymbol{b}) = \sum_{k=0}^{\infty} \sum_{\sum in_i = \sum ib_i - 2k} \widehat{\mathcal{N}}_{k,g,m-2k}(\boldsymbol{b}, \boldsymbol{n}) N^{n_0}\, s^{\sum ib_i - 2k}\, p^{\sum_{i\geq 1} n_i}, \tag{30}$$

where $\widehat{\mathcal{N}}_{k,g,m-2k}(\boldsymbol{b}, \boldsymbol{n})$ is the coefficient of $x^{2g-2} y^k s_0^{n_0} s_1^{n_1} \dots t_1^{b_1} t_2^{b_2} \dots$ in the power series expansion of e^F, $b = \sum_{i\geq 1} b_i$.

In [16], there is a matrix integral interpretation for the numbers $\mathcal{N}_{0,k,b}(\boldsymbol{p})$. Namely, let

$$\mathcal{F}(\boldsymbol{s}, t) = \sum_{\boldsymbol{p}} \sum_{b=1}^{\infty} \mathcal{N}_{0,k,b}(\boldsymbol{p}) \boldsymbol{s}^{\boldsymbol{p}} t^b,$$

then one has

$$\log \int_{\mathcal{H}(N)} \exp\left(-\frac{N}{2}TrX^2 + \sum_{k=1}^{\infty} \frac{t^k}{k} Tr(XA)^k\right) dX \to \mathcal{F}(\boldsymbol{s}, t),$$

where A is some matrix such that $s_i = \frac{1}{iN}\, TrA^i$.

For the large N limit in the 1-backbone case, this matrix integral can be modified so that the limit distribution is computable by free probability methods. Namely, consider semi-positive definite matrix $AXA^*AX^*A^*$. For any integer $k > 0$, we have

$$\lim_{N\to\infty} \frac{1}{N} \int Tr\left(AXA^*(AXA^*)^*\right)^k \exp\left(-\frac{N}{2}TrXX^*\right) dX = \sum_{\boldsymbol{p}} \mathcal{N}_{0,k,b}(\boldsymbol{p})\boldsymbol{s}^{\boldsymbol{p}}, \quad (31)$$

where $s_i = \frac{1}{N}Tr(AA^*)^i$.

Asymptotic spectral distribution. In the one backbone case, we can compute the leading term in the asymptotics of the matrix integral. To treat the large N limit of (27), one can use the techniques of free probability. Put $p = [qN]$ with some $q \in (0, 1)$ and consider the limit

$$\widetilde{M}_m(s, q) = \lim_{N\to\infty} N^{-m/2-1} M_m(s\sqrt{N}, p, N). \quad (32)$$

This limit is a polynomial in q and s, and the coefficient at $s^l q^n$ is the number of chord diagrams with $\frac{m-l}{2}$ chords, l marked points and n boundary components containing at least one marked point (i.e., $n = \sum_{i\geq 1} n_i$). Note that $\widetilde{M}_m(s, q)$ are the moments of a probability measure on $\mathbb{R}$, namely, the limit spectral distribution of the matrices $X/\sqrt{N} + sP$. This measure is uniquely determined by the limit spectral measures of $X/\sqrt{N}$ and sP.

Let us define the $\mathcal{R}$-transform $\mathcal{R}_\mu(z)$ and the $\mathcal{S}$-transform $\mathcal{S}_\mu(z)$ of a measure μ. We start with the moment generating function $\mathcal{M}_\mu(z)$ and the Cauchy transform $G_\mu(z)$ which are defined by the series

$$G_\mu(z) = \sum_{m=0}^{\infty} \frac{M_m}{z^{m+1}}, \quad (33)$$

and

$$\mathcal{M}_\mu(z) = \sum_{m=1}^{\infty} M_m z^m, \quad (34)$$

where $M_m = \int_{\mathbb{R}} x^m d\mu(x)$ are the moments of the measure μ. The (unique) solution of the equation

$$\mathcal{R}_\mu(G_\mu(z)) + \frac{1}{G_\mu(z)} = z. \quad (35)$$

is $\mathcal{R}_\mu(z)$. The $\mathcal{S}$-transform is defined by

$$\mathcal{S}_\mu(z) = \frac{z+1}{z}\mathcal{M}_\mu^{-1}(z). \quad (36)$$

The following is standard [24]:

Proposition 9. *(1) If A_N and B_N are two random Hermitian $N \times N$ matrices in general position, and the limit spectral distributions of A_N and B_N are μ and*

ν respectively, then the limit spectral distribution of $A_N + B_N$ is some distribution $\mu \boxplus \nu$, which is determined by its $\mathcal{R}$-transform

$$\mathcal{R}_{\mu\boxplus\nu}(z) = \mathcal{R}_\mu(z) + \mathcal{R}_\nu(z).$$

(2) If A_N and B_N are two random $N \times N$ matrices in general position, and the limit spectral distributions of $A_N A_N^*$ and $B_N B_N^*$ are μ and ν respectively, then the limit spectral distribution of $A_N B_N (A_N B_N)^*$ is some distribution $\mu \boxtimes \nu$, which is determined by its $\mathcal{S}$-transform

$$\mathcal{S}_{\mu\boxtimes\nu}(z) = \mathcal{S}_\mu(z)\mathcal{S}_\nu(z).$$

Thus, if one knows the $\mathcal{R}$-transform of the spectral distribution of $X/\sqrt{N}$ (let it be μ) and of sP (let it be ν), then one also knows the $\mathcal{R}$-transform of the spectral distribution of $X/\sqrt{N}+sP$. Computing the Cauchy transform of the latter and expanding it in the inverse powers of z, one gets the coefficients $\widetilde{M}_m(s,q)$ in accordance with (33). Note that the measure μ appears in the famous Wigner semicircle law, i.e.,

$$d\mu(x) = \begin{cases} \frac{1}{2\pi}\sqrt{4-x^2}, \text{ if } -2 \le x \le 2 \\ 0, \text{ otherwise,} \end{cases}$$

and the measure ν is a two-point distribution $q\delta(x-1) + (1-q)\delta(x)$. Now we compute the Cauchy transforms:

$$\begin{aligned} G_\mu(z) &= \frac{z - \sqrt{z^2-4}}{2}, \\ G_\nu(z) &= \frac{q}{z-s} + \frac{1-q}{z} = \frac{z-s+qs}{z(z-s)} \end{aligned}$$

and explicitly solve the equation (35):

$$\begin{aligned} \mathcal{R}_\mu(z) &= z, \\ \mathcal{R}_\nu(z) &= \frac{zs - 1 + \sqrt{(zs-1)^2 + 4zsq}}{2z}. \end{aligned}$$

Thus, the Cauchy transform $G(z)$ of the limit spectral measure of the matrix $X/\sqrt{N} + sP$ satisfies the equation

$$G(z) + s + \frac{\sqrt{(sG(z)-1)^2 - 4sqG(z)}}{2G(z)} = z.$$

This allows us to compute the first several polynomials $\widetilde{M}_m(s,q)$, and we find that they coincide with previous computations: $\widetilde{M}_m(s,q)$ is the coefficient of y^m found in Remark 2 by purely combinatorial methods.

Let us use the $\mathcal{S}$-transform technique to compute the limit spectral distribution of a random matrix $AXA^*AX^*A^*$, where X is a standard complex Gaussian $N \times N$ matrix with variance $N^{-1/2}$, and A is an $N \times N$ matrix such that $\lim_{N\to\infty} \frac{1}{N} Tr(AA^*)^k = s_k$. In other words, s_k are the moments of the limit spectral distribution of AA^* that we denote by ν. Its $\mathcal{S}$-transform is given by

$$\mathcal{S}_\nu(z) = \frac{z+1}{z}\,\mathcal{M}_\nu^{-1}(z)\;,$$

where

$$\mathcal{M}_\nu(z) = \sum_{k=1}^{\infty} s_k z^k$$

and $\mathcal{M}_\nu^{-1}$ is the inverse function to $\mathcal{M}_\nu$. The limit spectral distribution of XX^* (let it be μ) is the Marchenko–Pastur distribution with parameter 1 and has $\mathcal{S}$-transform of the form

$$\mathcal{S}_\mu(z) = \frac{1}{1+z}.$$

The limit spectral distribution λ of $AXA^*AX^*A^*$ therefore has $\mathcal{S}$-transform

$$\mathcal{S}_\lambda(z) = \frac{1}{1+z}\,\mathcal{S}_\nu^2(z).$$

The previous equation allows to compute length spectra for planar diagrams on one backbone, namely:

Theorem 3. *Put $\mathcal{K}(z) = \frac{z}{1+z}\mathcal{S}_\lambda(z)$. Then the one backbone generating function $G_1(x, z; \boldsymbol{s})$ for boundary length spectra in genus zero is given by*

$$G_1(0, z; \boldsymbol{s}) = 1 + \mathcal{K}^{-1}(z).$$

In particular, we have $G_1(0, z; 1, 1 \ldots) = C_{0,1}(z)$, the Catalan generating function.

The proof of this theorem immediately follows from (31), (36) and Proposition 9, (2). To check that $\mathcal{K}^{-1}(z)$ generates the Catalan numbers, we notice that for $s_k = 1$ for all k we have $\mathcal{M}_\nu(z) = \frac{z}{1-z}$ and $\mathcal{M}_\nu^{-1}(z) = \frac{z}{1+z}$. Therefore, $\mathcal{S}_\nu(z) = 1$ and $\mathcal{S}_\lambda(z) = \frac{1}{1+z}$. Thus we have $\mathcal{K}(z) = \frac{z}{(1+z)^2}$ and

$$\mathcal{K}^{-1}(z) = \frac{1 - 2z - \sqrt{1-4z}}{2z}$$

(since $\mathcal{K}^{-1}(0) = 0$), that is a well-known generating function for the Catalan numbers.

7. Closing Remarks

See [30, 31] for an application of the non-orientable diagrams to modeling the topology of proteins.

Inspired by the results of this paper and with an eye to understanding the multi-backbone analog of the differential equation equivalent to the Harer-Zagier recursion $(n+1)C_{g,1,n} = (2n-1)\left(2C_{g,1,n-1} + \binom{2n-2}{2}C_{g-1,1,n-2}\right)$, we ask if there is a differential operator in the variables (x, y, t) that vanishes on $C(x, y, t)$ and thus determines it. In fact, the Master Loop Equation of the model in [3] provides a constraint on $C(x, y, t)$ that however fails to give a differential operator.

Acknowledgements. We thank M. Kazarian for suggesting a link between Theorems 1, 2 and the KP theory.

References

[1] E . T. Akhmedov, S. Shakirov, Gluings of surfaces with polygonal boundaries, *Funct. Anal. Appl.*, **43** (2009), 245–253.

[2] J.E. Andersen, A. J. Bene, J.-B. Meilhan, R. C. Penner, Finite type invariants and fatgraphs, *Adv. Math.* **225** (2010), 2117–2161.

[3] J. E. Andersen, L. O. Chekhov, R. C. Penner, C. M. Reidys, P. Sulkowski, Topological recursion for chord diagrams, RNA complexes, and cells in moduli spaces, *Nucl. Phys. B*, **866** (2013), 414 – 443.

[4] J.E. Andersen, F.W.D. Huang, R.C. Penner, C.M. Reidys, Topology of RNA-RNA interaction structures, *Jour. Comp. Biol.*, **19** (2012), 928 – 943.
[5] J.E. Andersen, J. Mattes, N. Reshetikhin, The Poisson Structure on the Moduli Space of Flat Connections and Chord Diagrams, *Topology* **35** (1996), 1069–1083.
[6] J.E. Andersen, J. Mattes, N. Reshetikhin, Quantization of the algebra of chord diagrams, *Math. Proc. Camb. Phil. Soc.* **124** (1998), 451–467.
[7] J.E. Andersen, R.C. Penner, C.M. Reidys , M.S. Waterman, Topological classification and enumeration of RNA structures by genus, *Jour. Math. Biol.* DOI 10.1007/s00285-012-0594-x (2012), 1–18.
[8] J. E. Andersen, K. Ueno, Abelian conformal field theory and determinant bundles. Internat. J. Math. **18**, no. 8, (2007), 919–993.
[9] D. Bar-Natan, On the Vassiliev knot invariants. *Topology* **34** (1995), 423–475.
[10] M. Bon and H. Orland, TT2NE: A novel algorithm to predict RNA secondary structures with pseudoknots, *Nucl. Acids Res.* **41** (2010), 1895–1900.
[11] M. Bon, G. Vernizzi, H. Orland, A. Zee, Topological classification of RNA structures. *Jour. Mol. Biol.* **379** (2008), 900–911.
[12] R. Campoamor-Stursberg and V. O. Manturov, Invariant tensor formulas via chord diagrams, *Jour. Math. Sci.* **108** (2004), 3018–3029.
[13] O. Dumitsrescu, M. Mulase, A. Sorkin, B. Safnuk, The spectral curve of the Eynard-Orantin recursion via the Laplace transform, Contemp. Math. 593 (2013) 263-316.
[14] I. P. Goulden, D. M. Jackson, Transitive factorizations into transpositions and holomorphic mappings on the sphere, *Proc. Amer. Math. Soc.* **125** (1997), 5160.
[15] J. Harer and D. Zagier, The Euler characteristic of the moduli space of curves, *Inv. Math.* **85** (1986), 457–485
[16] V.A. Kazakov, M. Staudacher, T. Wynter, Character expansion methods for matrix models of dually weighted graphs, *Comm. Math. Phys.* **177** (1996) 451–468.
[17] M. Kazarian, KP hierarchy for Hodge integrals, *Adv. Math.* **221**, 1–21 (2009).
[18] M. Kazarian, *Private communication* (2013).
[19] K. Kontsevich, Vassiliev's knot invariants, *Adv. Sov. Math.* **16** (1993), 137–150.
[20] S. K. Lando, A. K. Zvonkin, Graphs on surfaces and their applications, *Encyclopaedia of Mathematical Sciences* **141**, Springer-Verlag, Berlin (2004).
[21] M. Ledoux, A recursion formula for the moments of the Gaussian orthogonal ensemble, *Ann. Inst. H. Poincare Prob. Stat.* **45** (2009), 754–769.
[22] T. Miwa, M. Jimbo, E. Date, Solitons: Differential equations, symmetries and infinite-dimensional algebras, *Cambridge Tracts in Mathematics* **135**, Cambridge University Press, Cambridge (2000).
[23] A. Morozov, Sh. Shakirov, Generation of matrix models by $\hat{W}$-operators, JHEP04(2009)064.
[24] A. Nica, R. Speicher, Lectures on the combinatorics of free probability, *London Math. Soc. Lecture Notes Series* **335** Cambridge University Press (2006).
[25] H. Orland and A. Zee, RNA folding and large N matrix theory, *Nucl. Phys. B* **620** (2002), 456-476.
[26] A. V. Pastor, O. P. Rodionova, Some formulas for the number of gluings, *Zapiski Nauchnykh Seminarov POMI* **406** (2012), 117–156.
[27] R. C. Penner, The decorated Teichmuller space of a punctured surface, *Comm. Math. Phys.* **113** (1987), 299–339.
[28] R. C. Penner and M. S. Waterman, Spaces of RNA secondary structures, *Adv. Math.* **101** (1993), 31–49.
[29] R. C. Penner, Cell decomposition and compactification of Riemann's moduli space in decorated Teichmueller theory. In: N. Tongring and R. C. Penner (eds) Woodshole mathematics-perspectives in math and physics. World Scientific, Singapore, 263–301 (2004).
[30] R. C. Penner, M. Knudsen, C. Wiu, J. E. Andersen, Fatgraph model of proteins, *Comm. Pure Appl. Math.* **63** (2010), 1249–1297.
[31] R.C. Penner, M. Knudsen, C. Wiuf, J. E. Andersen, An Algebro-Topological Model of Protein Domain Structure, *PLoS ONE* **6** (2011), 1 – 14.
[32] M. Pillsbury, H. Orland, A. Zee, Steepest descent calculation of RNA pseudoknots, *Phys. Rev. E* **72** (2010), 011911.
[33] M. Pillsbury, J. A. Taylor, H. Orland, A. Zee, An algorithm for RNA pseudoknots, arXiv: cond-mat/0310505v2 (2005).

[34] C. M. Reidys, Combinatorial and Computational Biology of Pseudoknot RNA, *Applied Math series,* Springer (2010).

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

[36] K. Strebel, Quadratic differentials, Ergebnisse der Mathematik und ihrer Grenzgebiete (3), 5. Springer-Verlag, Berlin (1984).

[37] G. Vernizzi, H. Orland, A. Zee, Enumeration of RNA structures by matrix models. *Phys. Rev. Lett.* **94** (2005), 168103.

[38] G. Vernizzi, P. Ribecca, H. Orland, A. Zee, Topology of pseudoknotted homopolymers, *Phys. Rev. E* **73** (2006), 031902.

Chebyshev Laboratory, St. Petersburg State University, 14th Line V.O. 29B, St. Petersburg 199178 Russia

E-mail address: nikita.v.alexeev@gmail.com

Centre for Quantum Geometry of Moduli Spaces, Department of Mathematics, University of Aarhus, DK-8000 Aarhus C, Denmark

E-mail address: andersen@qgm.au.dk

Centre for Quantum Geometry of Moduli Spaces, Department of Mathematics, University of Aarhus, DK-8000 Aarhus C, Denmark, and Mathematics Department, Caltech, Pasadena, CA 91125 USA

E-mail address: rpenner@caltech.edu

St.Petersburg Department, Steklov Mathematical Institute, Fontanka 27, St. Petersburg 191023, and Chebyshev Laboratory, St. Petersburg State University, 14th Line V.O. 29B, St.Petersburg 199178 Russia

E-mail address: zograf@pdmi.ras.ru