MONOTONE HURWITZ NUMBERS AND THE HCIZ INTEGRAL I

I. P. GOULDEN, M. GUAY-PAQUET, AND J. NOVAK

Abstract. We prove that the free energy of the Harish-Chandra-Itzykson-Zuber matrix model admits an $N \to \infty$ asymptotic expansion in powers of N^{-2} whose coefficients are generating functions for a desymmetrized version of the double Hurwitz numbers, which we call monotone double Hurwitz numbers. Thus, the HCIZ free energy expands as a generating function enumerating certain branched covers of the Riemann sphere with arbitrary branching over 0 and ∞ and simple branching elsewhere. We prove that the monotone double Hurwitz numbers exhibit the main structural properties of the usual double Hurwitz numbers: their total generating function is a solution of the 2D Toda Lattice equations, and the numbers themselves are piecewise polynomial functions on pairs of partitions.

Contents

Date: July 7, 2011.
1991 *Mathematics Subject Classification.* Primary 05A15, 14E20; Secondary 15B52.
Key words and phrases. Hurwitz numbers, matrix models, enumerative geometry.
IPG and MG-P were supported by NSERC.

0. Introduction

The Harish-Chandra-Itzykson-Zuber matrix model is a complex, unit-mass Borel measure μ_N on the group of $N \times N$ unitary matrices. This measure is by definition absolutely continuous with respect to the Haar probability measure on $\mathbf{U}(N)$, being given by the density

$$\mu_N(\mathrm{d}U) = \frac{1}{I_N(z)} e^{zN\,\mathrm{tr}(AUBU^*)}\mathrm{d}U, \tag{0.1}$$

where z is a complex parameter and A, B are $N \times N$ complex matrices. The partition function of the model,

$$I_N(z) = \int e^{zN\,\mathrm{tr}(AUBU^*)}\mathrm{d}U, \tag{0.2}$$

is known as the *Harish-Chandra-Itzykson-Zuber integral*. This integral was first considered by Harish-Chandra in his study of differential operators on semisimple Lie algebras [30], where it was evaluated as a ratio of determinants under the assumption that A, B lie in the Lie algebra of the unitary group. It appears in various other contexts, including the study of large deviations in the spectral measure of Gaussian sample covariance matrices [28], as the reproducing kernel of a distinguished inner product on symmetric polynomials [48], and especially in the theory of matrix models, where Harish-Chandra's formula was independently rediscovered by Itzykson and Zuber in the course of their study of the Hermitian two-matrix model [34, 67]. It is in this latter context that the problem of analyzing the $N \to \infty$ asymptotic behaviour of the free energy

$$F_N(z) = \frac{1}{N^2} \log I_N(z) \tag{0.3}$$

of the HCIZ model first arose. This problem has since received considerable attention from mathematicians, notably in the work of Guionnet and her collaborators, who have addressed it using the methods of large deviation theory, Schwinger-Dyson equations, and non-commutative differential calculus [10, 25, 26, 28].

Apart from its importance in the asymptotic analysis of Hermitian multi-matrix models, which is reviewed in [26, 27], the large N behaviour of the HCIZ free energy is of interest for another, deeper reason. It is well-known that the free energy of Hermitian matrix models admits a perturbative expansion in powers of N^{-2}, often called the "genus expansion," each order of which is a generating function enumerating polygonal discretizations of a compact two-dimensional surface of given topology. The genus expansion is a key ingredient in the formulation of Witten's famous conjecture relating two-dimensional quantum gravity to intersection theory on the moduli space of curves [64]. Several proofs of Witten's conjecture have been given [37, 39, 44, 49]. Theoretical physicists have postulated that the free energy of the HCIZ matrix model admits an analogous genus expansion with similar ties to algebraic combinatorics and enumerative algebraic geometry [34, 43, 54]. In this article and its sequel [18], we provide a precise formulation and proof of this conjecture.

0.1. **Main results and organization.** Our main result concerning the asymptotics of the HCIZ model is the following.

Theorem 0.1. *Let $(A_N), (B_N)$ be two sequences of $N \times N$ normal matrices whose spectral radii are uniformly bounded, with least upper bound*

$$M := \sup \{\rho(A_N), \rho(B_N) : N \geq 1\},$$

and which admit limiting moments

$$-\phi_k := \lim_{N\to\infty} \frac{1}{N} \operatorname{tr}(A_N^k)$$
$$-\psi_k := \lim_{N\to\infty} \frac{1}{N} \operatorname{tr}(B_N^k)$$

of all orders. Let $0 \leq r < r_c$, where r_c is the critical value

$$r_c = \frac{2}{27}.$$

Then, the free energy $F_N(z)$ admits an $N \to \infty$ asymptotic expansion of the form

$$F_N(z) \sim \sum_{g=0}^{\infty} \frac{C_g(z)}{N^{2g}}$$

which holds uniformly on the closed disc $\overline{D}(0, rM^{-2})$. Each coefficient $C_g(z)$ is a holomorphic function of z on the open disc $D(0, r_cM^{-2})$, with Maclaurin series

$$C_g(z) = \sum_{d=1}^{\infty} C_{g,d} \frac{z^d}{d!},$$

where

$$C_{g,d} = \sum_{\alpha,\beta\vdash d} \vec{H}_g(\alpha, \beta)\phi_\alpha\psi_\beta$$

and $\vec{H}_g(\alpha, \beta)$ is the number of $(r + 2)$-tuples $(\sigma, \rho, \tau_1, \ldots, \tau_r)$ of permutations from the symmetric group $\mathbf{S}(d)$ such that

(1) *σ has cycle type α, ρ has cycle type β, and the τ_i are transpositions;*
(2) *The product $\sigma\rho\tau_1 \ldots \tau_r$ equals the identity permutation;*
(3) *The group $\langle\sigma, \rho, \tau_1, \ldots, \tau_r\rangle$ acts transitively on $\{1, \ldots, d\}$;*
(4) *$r = 2g - 2 + \ell(\alpha) + \ell(\beta)$;*
(5) *Writing $\tau_i = (s_i\ t_i)$ with $s_i < t_i$, we have $t_1 \leq \cdots \leq t_r$.*

Theorem 0.1 is proved in Section 1 below. It should be considered as the HCIZ analogue of the genus expansion for Hermitian matrix models. The genus expansion of the Hermitian one-matrix model was conjectured in the physics literature by Bessis, Itzykson, and Zuber [4], building on insights of 't Hooft [61], and later rigorously proved by Ercolani and MacLaughlin [17] and Bleher and Its [6], see also the recent work of Bleher and Deaño [5]. Convergence of the free energy of Hermitian multi-matrix models to generating functions for coloured planar maps was established by Guionnet in [25].

The geometric meaning of the asymptotic expansion claimed in Theorem 0.1, which justifies its designation as a "genus expansion," is best understood by pursuing the analogy with Hermitian matrix models. The genus expansion of the Hermitian one-matrix model encodes a combinatorial method for constructing a compact Riemann surface (or smooth projective curve), namely by glueing together polygonal tiles cut out of the complex plane. Another recipe for constructing a compact Riemann surface is to realize it as a branched covering of the Riemann sphere (or projective line) $\mathbb{P}^1$. Indeed, it is a consequence of the Riemann Existence Theorem that, given a finite-sheeted topological branched covering $f : S \to \mathbb{P}^1$ of the Riemann sphere by a compact surface S, there is a unique complex structure on S which makes this map holomorphic. A classical construction due to Hurwitz [32, 33] encodes a given d-sheeted branched covering as a transitive factorization of the identity in $\mathbf{S}(d)$. Such factorizations are often called *constellations* [41]. Hurwitz's construction takes as input a branched covering $f : S \to \mathbb{P}^1$ together with a labelling of the branch points of f and a labelling of the points in the fibre of f over a specified unbranched basepoint, and outputs a constellation whose factors are determined by the lifts of small loops on the sphere encircling the branch points of f. The cycle type of each factor coincides with the monodromy of f over the corresponding branch point.

A particular case of Hurwitz's construction produces the *double Hurwitz numbers* $H_g(\alpha, \beta)$ considered by Okounkov [46] and further studied by Goulden, Jackson, and Vakil [23]. Given two partitions $\alpha, \beta \vdash d$, the double Hurwitz number counts (up to an appropriate notion of isomorphism) degree d branched covers $f : S \to \mathbb{P}^1$ of the Riemann sphere in which the source curve S has genus g and the map f has monodromy α over 0, β over ∞, and $r = 2g - 2 + \ell(\alpha) + \ell(\beta)$ additional simple branch points at fixed positions, the number of which is determined by the Riemann-Hurwitz formula. Applying Hurwitz's construction, $H_g(\alpha, \beta)$ counts[1] $(r+2)$-tuples $(\sigma, \rho, \tau_1, \ldots, \tau_r)$ of permutations from $\mathbf{S}(d)$ verifying the first four of the combinatorial conditions listed in Theorem 0.1.

Thus, the coefficients $C_g(z)$ in the asymptotic expansion of the HCIZ free energy are generating functions enumerating certain branched covers of $\mathbb{P}^1$ by curves of genus g with arbitrary ramification over two given points and simple branching over an appropriate number of additional fixed points. The fifth combinatorial condition in Theorem 0.1, which distinguishes the *monotone double Hurwitz numbers* $\vec{H}_g(\alpha, \beta)$ from Okounkov's double Hurwitz numbers, is a special feature of the HCIZ model whose origin will become clear below. In a sense, it is an elaboration of the relationship between the Catalan numbers $\frac{1}{d+1}\binom{2d}{d}$, which count monotone trees, monotone parking functions, etc., and the Cayley numbers d^{d-2}, which count symmetrized structures of the same type. This analogy is developed in full in our second paper on this subject [18], which contains a detailed combinatorial analysis of the *single* monotone Hurwitz numbers $\vec{H}_g(\alpha) = \vec{H}_g(\alpha, (1^d))$ proceeding in tandem with the well-developed combinatorial theory of the single Hurwitz numbers $H_g(\alpha) = H_g(\alpha, (1^d))$.

In connection with two-dimensional quantum chromodynamics, physicists have come to consider the enumeration of branched covers of $\mathbb{P}^1$ as a closed string theory

[1]The usual definition of the Hurwitz numbers would include a further division by $d!$ in order to compensate for reparameterizations of the domain. For our purposes, it will be more convenient to omit this division — this is like working with labelled maps instead of unlabelled maps.

on the Riemann sphere [24]. According to Theorem 0.1, the free energy of the HCIZ model approximates the partition function of such a theory under a combinatorial constraint. One may study the limit object directly once it has been identified. Section 2 contains our second set of results, which focus on the structure of the monotone double Hurwitz numbers $\vec{H}_g(\alpha,\beta)$. The usual double Hurwitz numbers enjoy several remarkable properties, the most important of which are the connection with integrable hierarchies of partial differential equations [46, 53] and piecewise polynomial dependence on ramification type when the genus and number of ramification points over 0 and ∞ are held fixed [23, 35, 57]. The goal of Section 2 is to prove that the monotone double Hurwitz numbers exhibit these properties.

Theorem 0.2. *Let z, q be indeterminates, and let $A = \{a_1, a_2, \dots\}, B = \{b_1, b_2, \dots\}$ be two alphabets of indeterminates. The formal power series*

$$\vec{\mathbf{H}}(z,q,A,B) = \sum_{d=1}^{\infty} \frac{z^d}{d!} \sum_{r=0}^{\infty} q^r \sum_{\alpha,\beta \vdash d} \vec{H}^r(\alpha,\beta) p_\alpha(A) p_\beta(B)$$

is a solution of the 2D Toda lattice hierarchy in the two sets of variables $p_1(A), p_2(A), \dots,$ $p_1(B), p_2(B), \dots,$ where

$$\begin{array}{ll} p_1(A) = a_1 + a_2 + \dots & p_1(B) = b_1 + b_2 + \dots \\ p_2(A) = a_1^2 + a_2^2 + \dots & p_2(B) = b_1^2 + b_2^2 + \dots \\ \vdots & \vdots \end{array}$$

are the power-sum symmetric functions in the alphabets A and B.

This result establishes the monotone analogue of the main result of [46]. The equations of the Toda hierarchy yield a countable set of recurrences which uniquely determine the monotone double Hurwitz numbers. Furthermore, via Theorem 0.1, every result about the monotone double Hurwitz numbers also furnishes information on the HCIZ model. Theorem 0.2 describes how the asymptotic expansion of the HCIZ free energy evolves as the limiting moments of the matrices A_N and B_N vary. In particular, Theorem 0.2 generalizes a result of Zinn-Justin [66], who proved that in the large N limit the free energy F_N becomes a solution of a limiting version of the 2D Toda hierarchy (the *dispersionless* 2D Toda hierarchy).

Finally, we prove that the monotone double Hurwitz numbers $\vec{H}_g(\alpha,\beta) = \vec{H}_g(\alpha_1,\dots,\alpha_m,\beta_1,\dots,\beta_n)$ are piecewise polynomial functions of α_i, β_j for fixed g, m, n. The piecewise polynomiality of the usual double Hurwitz numbers $H_g(\alpha,\beta)$ was proved by Goulden, Jackson and Vakil [23] and serves to support their conjecture that the double Hurwitz numbers are top intersections on some universal Picard variety. It is therefore desirable to formulate an analogous statement for the monotone double Hurwitz numbers.

Theorem 0.3. *To each triple (g, m, n) consisting of a non-negative integer g and positive integers m, n, there corresponds a hyperplane arrangement in $\mathbb{R}^{m+n}$ and a collection of polynomials $\vec{P}_{g,\mathfrak{c}}(x_1,\dots,x_m,y_1,\dots,y_n)$ indexed by the chambers of this hyperplane arrangement such that*

$$\frac{|\operatorname{Aut}\alpha||\operatorname{Aut}\beta|}{|\alpha|!}\vec{H}_g(\alpha,\beta) = \vec{P}_{g,\mathfrak{c}}(\alpha_1,\dots,\alpha_m,\beta_1,\dots,\beta_n)$$

for all partitions α, β with $|\alpha| = |\beta|, \ell(\alpha) = m, \ell(\beta) = n,$ and $(\alpha,\beta) \in \mathfrak{c}$.

We will see below that the hyperplane arrangement which determines the piecewise polynomiality of the monotone double Hurwitz numbers is precisely the *resonance arrangement* of [35, 56], which determines the piecewise polynomiality of the usual Hurwitz numbers.

0.2. **Some remarks on our second paper.** Let us close this Introduction with a brief discussion of our second paper on this subject [18].

Loosely speaking, the HCIZ matrix model is "like" the Hermitian two-matrix model: its free energy is a solution of the 2D Toda hierarchy that counts combinatorial/geometric structures with two degrees of enumerative freedom. When one of the matrix sequences defining the HCIZ potential, say (B_N), has degenerate limiting moments $-\psi_k = \delta_{1k}$, Theorem 0.1 tells us that the HCIZ model degenerates from monotone double Hurwitz theory to monotone single Hurwitz theory. It should come as no surprise to the reader familiar with Hurwitz theory that the monotone single Hurwitz numbers are much more accessible than their double counterparts: we have explicit formulas for $\vec{H}_g(\alpha)$ in genus $g = 0, 1$ [18, Theorems 0.3 and 0.4], and we are able to prove directly that an ELSV-type polynomiality property holds in all genera, thus establishing the general form of the single monotone Hurwitz numbers.

From the point of view of random matrix theory, these results show that the degenerate HCIZ model is "like" the Hermitian one-matrix model: its free energy is a solution of the KP hierarchy, we can solve explicitly for the leading and sub-leading orders, and for genus two and higher we find that all $C_g(z)$ are rational functions of a single explicit algebraic function of z (the analogous property for the Hermitian one-matrix model was conjectured in [4] and has been proved very recently by Ercolani [16]). Our paper [18] gives a detailed combinatorial treatment of single monotone Hurwitz theory leading to the counterparts of most key results in classical single Hurwitz theory, thereby providing an essentially complete understanding of the "one-sided" HCIZ model.

0.3. **Acknowledgements.** It is a pleasure to acknowledge helpful conversations with our colleagues Sean Carrell, David Jackson and Bruce Richmond at the University of Waterloo, as well as correspondence with Jamie Mingo of Queen's University.

J. N. would like to thank the organizers of the Fall 2010 MSRI semester program "Random Matrix Theory, Interacting Particle Systems and Integrable Systems" for the opportunity to participate as a postdoctoral fellow. He gratefully acknowledges helpful interaction with the other program participants, especially Percy Deift, Peter Forrester and Ken McLaughlin.

1. Asymptotic expansion of the free energy

The HCIZ model is analytically simpler and combinatorially more involved than Hermitian matrix models. Just as the Hermitian one-matrix model is a deformation of Wigner's Gaussian Unitary Ensemble, the HCIZ model is a deformation of Dyson's Circular Unitary Ensemble. Unlike the partition function in Hermitian matrix model theory, the partition function of the HCIZ model is entire. Thus we can in principle access the asymptotics of $F_N(z)$ through a clear understanding of the asymptotics of its Taylor coefficients. Unfortunately, the asymptotic behaviour of the Taylor coefficients is not so easily obtained as in the Hermitian case, where one has the machinery of Wick calculus. Following [9, 11, 42, 45], we use the invariant theory of the unitary group to derive a Wick-type Lemma for the integration of polynomial functions on $\mathbf{U}(N)$. The $\mathbf{U}(N)$-Wick Lemma is then used to obtain the full $N \to \infty$ asymptotic expansion of the Taylor coefficients $F_N^{(d)}(0) = \frac{\partial^d}{\partial z^d} F_N(z)|_{z=0}$ for each positive integer d. This expansion takes the form

$$F_N^{(d)}(0) \sim \sum_{g=0}^{\infty} \frac{C_{g,d}}{N^{2g}}, \tag{1.1}$$

where the hypotheses and notation are those of Theorem 0.1. The proof of Theorem 0.1 is then reduced to verifying that the generating functions

$$C_g(z) = \sum_{d=1}^{\infty} C_{g,d} \frac{z^d}{d!} \tag{1.2}$$

all converge in the disc $D(0, r_c M^{-2})$, where $r_c = 2/27$ is the critical value claimed in Theorem 0.1. This is done by reducing the problem to that of determining the radius of convergence of the generating functions

$$\vec{\mathbf{H}}_g^s(z) = \sum_{d=1}^{\infty} \vec{H}_{g,d} \frac{z^d}{d!}, \tag{1.3}$$

where $\vec{H}_{g,d} = \vec{H}_g(1^d, 1^d)$ is the monotone *simple* Hurwitz number enumerating constellations which consist of monotone transpositions only. The common singularity of the generating functions $\vec{\mathbf{H}}_g^s(z)$ is detected by specializing the results of our second paper [18].

1.1. **Basic properties of the HCIZ model.** Since the HCIZ model is unitarily invariant, if we assume that the matrices A, B which define its density are normal then we may also assume that they are diagonal without further loss in generality:

$$\begin{aligned} A &= \operatorname{diag}(a_1, \dots, a_N) \\ B &= \operatorname{diag}(b_1, \dots, b_N). \end{aligned} \tag{1.4}$$

This assumption will be made from now on.

Proposition 1.1. *For any $z \in \mathbb{C}$, we have*

$$|I_N(z)| \leq e^{N^2 \rho(A)\rho(B)|z|},$$

where

$$\begin{aligned} \rho(A) &= \max_{1 \leq i \leq N} |a_i| \\ \rho(B) &= \max_{1 \leq j \leq N} |b_j| \end{aligned}$$

are the spectral radii of A and B.

Proof. Since the Haar measure is a positive Borel measure, we have

$$|I_N(z)| = \left| \int e^{zN \operatorname{tr}(AUBU^*)} \mathrm{d}U \right| \leq \int \left| e^{zN \operatorname{tr}(AUBU^*)} \right| \mathrm{d}U \leq \int e^{|zN \operatorname{tr}(AUBU^*)|} \mathrm{d}U.$$

The result now follows since

$$\left| \operatorname{tr} AUBU^* \right| = \left| \sum_{i=1}^N \sum_{j=1}^N a_i b_j u_{ij} \overline{u}_{ij} \right| \leq \sum_{i=1}^N \sum_{j=1}^N |a_i||b_j||u_{ij}|^2 \leq N\rho(A)\rho(B),$$

where we have used the fact that the columns of a unitary matrix are unit vectors. □

Proposition 1.2. *$I_N(z)$ is an entire function of $z \in \mathbb{C}$, with Maclaurin series*

$$I_N(z) = \sum_{d=0}^\infty \left(N^d \int (\operatorname{tr} AUBU^*)^d \mathrm{d}U \right) \frac{z^d}{d!}.$$

Proof. We will prove that the derivative of $I_N(z)$ can be computed at any point $z \in \mathbb{C}$ by differentiating under the integral sign:

$$I_N'(z) = \int \frac{\partial}{\partial z} e^{zN \operatorname{tr}(AUBU^*)} \mathrm{d}U = N \int (\operatorname{tr} AUBU^*) e^{zN \operatorname{tr}(AUBU^*)} \mathrm{d}U.$$

The Taylor expansion of $I_N(z)$ is obtained by repeated differentiation under the integral sign.

We have

$$I_N(z) = \int K_N(z, U) \mathrm{d}U,$$

where the kernel $K_N(z,U) = e^{zN \operatorname{tr}(AUBU^*)}$ is an entire function of $z \in \mathbb{C}$ and a continuous function of $(z,U) \in \mathbb{C} \times \mathbf{U}(N)$. Let $z \in \mathbb{C}$ be given, and let $(z_n)_{n=1}^\infty$ be a sequence of complex numbers, each distinct from z, which converge to z. Then $\{z_n\} \times \mathbf{U}(N)$ is a compact set and the Newton quotient

$$\frac{K_N(z_n, U) - K_N(z, U)}{z_n - z}$$

is a continuous, hence bounded, function on this set. Thus we may apply the dominated convergence theorem:

$$\lim_{n \to \infty} \frac{I_N(z_n) - I_N(z)}{z_n - z} = \int \lim_{n \to \infty} \frac{K_N(z_n, U) - K_N(z, U)}{z_n - z} \mathrm{d}U = \int \frac{\partial}{\partial z} K_N(z, U) \mathrm{d}U.$$

□

Propositions 1.1 and 1.2 together say that $I_N(z)$ is an entire function of order 1. Since $I_N(0) = 1$, the partition function is non-vanishing in a neighbourhood of $z = 0$, and we define the free energy $F_N(z)$ as the unique holomorphic function on this neighbourhood such that

$$e^{N^2 F_N(z)} = I_N(z), \quad F_N(0) = 0. \tag{1.5}$$

1.2. **Matrix group Wick Lemma.** Let $G \subseteq \mathbf{U}(N)$ be a closed subgroup of the unitary group. A generic element of G will be denoted $U = (u_{ij})$.

A function $f : G \to \mathbb{C}$ is said to be polynomial if there exists a polynomial p_f in N^2 variables such that

$$f(U) = p_f(u_{11}, \dots, u_{NN}) \tag{1.6}$$

for all $U \in G$. Let $\mathcal{A} \subset L^2(G, \mathrm{Haar})$ be the algebra of polynomial functions on G. This algebra admits the orthogonal decomposition

$$\mathcal{A} = \bigoplus_{d=0}^{\infty} \mathcal{A}^{(d)}, \tag{1.7}$$

where $\mathcal{A}^{(d)}$ is the space of homogeneous polynomial functions of degree d. Since the decomposition (1.7) is orthogonal, the computation of inner products $\langle f, g\rangle$ in $\mathcal{A}$ reduces to the case where f, g belong to the same space $\mathcal{A}^{(d)}$. Furthermore, by linearity of the integral, it suffices to consider the case where f, g are monomials:

$$\begin{aligned} &\langle u_{i(1)j(1)} \dots u_{i(d)j(d)}, u_{i'(1)j'(1)} \dots u_{i'(d)j'(d)} \rangle \\ &= \int u_{i(1)j(1)} \dots u_{i(d)j(d)} \overline{u_{i'(1)j'(1)} \dots u_{i'(d)j'(d)}} \mathrm{d}U. \end{aligned} \tag{1.8}$$

Such inner products of monomials will be called $(d+d)$-point correlation functions.

An analogue of the usual Gaussian Wick Lemma (see e.g. [41]) would be an algorithm which reduces the computation of $(d+d)$-point functions to the computation of simpler correlation functions. It is not immediately clear how to obtain such an algorithm, since there is apparently no analogue of the “propagator” in this setting. One may manufacture a propagator using the invariant theory of G. This gives an analogue of Wick's Lemma for G. Unfortunately, the G-Wick Lemma is less effective than the vector space Wick Lemma: it only reduces the computation of $(d+d)$-point functions to the computation of a distinguished subclass of $(d+d)$-point functions, not to $(1+1)$-point functions (covariances). Moreover, in order for the G-Wick Lemma to be practically useful, one must have available a complete description of the G-invariants in the full mixed tensor algebra over the defining representation of G. Following [9, 11, 42, 45], we will indicate how the argument proceeds in general, and restrict to the case $G = \mathbf{U}(N)$ when necessary.

Let V be the defining representation of G with its standard inner product, and let V^* be the dual representation. Let $\{e_1, \dots, e_N\}$ be an orthonormal basis of V. Consider the G-module $V_{dd} = V^{\otimes d} \otimes (V^*)^{\otimes d}$ of mixed tensors of type (d, d).

Let P be the matrix of the orthogonal projection $V_{dd} \to V_{dd}^G$ of the space of mixed tensors onto the subspace of G-invariant tensors, with respect to the standard basis

$$e_{i(1)} \otimes \dots \otimes e_{i(d)} \otimes e^*_{j(1)} \otimes \dots \otimes e^*_{j(d)}, \quad i, j : [d] \to [N] \tag{1.9}$$

of V_{dd}. Since G is a compact group, the orthogonal projection $V_{dd} \to V_{dd}^G$ may be obtained by averaging the action of G against the Haar probability measure. Consequently, the matrix elements of P are precisely the $(d+d)$-point functions (1.8). Now choose a basis of the invariant subspace V_{dd}^G, and let A be the $\dim V_{dd} \times \dim V_{dd}^G$ rectangular matrix whose columns are the coordinates of these basis vectors with respect to the basis (1.9). Then, since P is an orthogonal projection, it factors as

$$P = A(A^*A)^{-1}A^* \tag{1.10}$$

(this is the "outer product divided by inner product" formula for orthogonal projections familiar from linear algebra, see e.g. [59]).

We can extract a Wick-type formula from this factorization by taking a general matrix element of P on the left, and equating it with the corresponding matrix element on the right obtained simply from the definition of matrix multiplication. This will give a formula for the general $(d+d)$-point function in terms of the matrix elements of A and Γ^{-1}, where $\Gamma = A^*A$. Thus, in a sense, the inverse Gram matrix Γ^{-1} is playing the same role as that played by the propagator (or covariance matrix) in the Gaussian vector space Wick Lemma.

In order for the above general reasoning to produce useful formulas, we must have a complete understanding of the space of tensor invariants V_{dd}^G. Let us now restrict to the case $G = \mathbf{U}(N)$, where we have an explicit basis for the space of tensor invariants.

We recall that a permutation $\pi \in \mathbf{S}(d)$ is said to have a *decreasing subsequence* of length k if $\pi(i_1) > \cdots > \pi(i_k)$ for some indices $1 \leq i_1 < \cdots < i_k \leq d$. For a survey of the combinatorics of increasing and decreasing subsequences in permutations, the reader is referred to [58]; here we only need this definition in order to state the following result of Baik and Rains.

Theorem 1.3 ([1, 31])**.** *Let $\mathbf{S}_N(d)$ denote the set of permutations in the symmetric group $\mathbf{S}(d)$ which have no decreasing subsequence of length $N+1$. Then the tensors*

$$t_\sigma = \sum_{i:[d]\to[N]} e_{i(\sigma(1))} \otimes \cdots \otimes e_{i(\sigma(d))} \otimes e^*_{i(1)} \otimes \cdots \otimes e^*_{i(d)}, \quad \sigma \in \mathbf{S}_N(d)$$

are linearly independent and span the space of invariant tensors V_{dd}^G.

This leads to the following $\mathbf{U}(N)$-Wick Lemma.

Lemma 1.4 ($\mathbf{U}(N)$-Wick Lemma)**.** *For any $i, j, i', j' : [d] \to [N]$, we have that*

$$\langle u_{i(1)j(1)} \dots u_{i(d)j(d)}, u_{i'(1)j'(1)} \dots u_{i'(d)j'(d)} \rangle = \sum_{\sigma,\rho} (\Gamma^{-1})_{\sigma\rho},$$

where the sum runs over pairs of permutations σ, ρ from the set $\mathbf{S}_N(d)$ such that $i = i' \circ \sigma$, $j = j' \circ \rho$, and $(\Gamma^{-1})_{\sigma\rho}$ is the (σ,ρ)-element of the inverse of the matrix

$$\Gamma = \begin{bmatrix} & \vdots & \\ \dots & N^{c(\sigma\rho^{-1})} \dots & \\ & \vdots & \end{bmatrix}_{\sigma,\rho \in \mathbf{S}_N(d)},$$

where $c(\pi)$ denotes the number of cycles in the permutation $\pi \in \mathbf{S}(d)$.

The $\mathbf{U}(N)$-Wick Lemma was first stated in the physics literature by Samuel [55]; it was rediscovered and proved by Collins in [9]. The general nature of the argument allows it to be extended to other classical groups [11], and even to the setting of compact matrix quantum groups in the sense of Woronowicz, see [2].

These developments have led to a statistical theory of random matrices sampled from compact quantum groups, see [3]. For our purposes, however, we will need to develop a finer understanding of the original $\mathbf{U}(N)$ case.

Applying the $\mathbf{U}(N)$-Wick Lemma to the Maclaurin series of the HCIZ integral, we obtain the following.

Proposition 1.5. *The Maclaurin series of $I_N(z)$ is*

$$I_N(z) = \sum_{d=0}^{\infty} \left(N^d \sum_{\sigma \in \mathbf{S}_N(d)} \sum_{\rho \in \mathbf{S}_N(d)} (\Gamma^{-1})_{\sigma\rho} p_\sigma(A) p_\rho(B) \right) \frac{z^d}{d!}$$

where, for σ of cycle type α, $p_\sigma(A)$ denotes the power-sum symmetric function p_α specialized at the eigenvalues of A (and similarly for $p_\rho(B)$).

Proof. Recall that our matrices A and B are diagonal: $A = \operatorname{diag}(a_1, \dots, a_N), B = \operatorname{diag}(b_1, \dots, b_N)$. We have

$$\operatorname{tr} AUBU^* = \sum_{i=1}^{N} \sum_{j=1}^{N} a_i b_j u_{ij} \overline{u}_{ij},$$

and more generally

$$(\operatorname{tr} AUBU^*)^d = \sum_{i:[d]\to[N]} \sum_{j:[d]\to[N]} a_{i(1)} \dots a_{i(d)} b_{j(1)} \dots b_{j(d)} u_{i(1)j(1)} \dots u_{i(d)j(d)} \overline{u_{i(1)j(1)} \dots u_{i(d)j(d)}}.$$

Applying the $\mathbf{U}(N)$-Wick Lemma to this expansion, we have

$$\int u_{i(1)j(1)} \dots u_{i(d)j(d)} \overline{u_{i(1)j(1)} \dots u_{i(d)j(d)}} \mathrm{d}U = \sum_{\substack{\sigma \in \mathbf{S}_N(d) \\ \sigma \in \operatorname{Aut}(i)}} \sum_{\substack{\rho \in \mathbf{S}_N(d) \\ \rho \in \operatorname{Aut}(j)}} (\Gamma^{-1})_{\sigma\rho},$$

where $\operatorname{Aut}(i)$ denotes the group of automorphisms of i under the natural action of $\mathbf{S}(d)$ on functions $i : [d] \to [N]$ by permutation of arguments. Thus

$$\begin{aligned} &I_N(z) \\ &= \sum_{d=0}^{\infty} \frac{z^d}{d!} N^d \sum_{i:[d]\to[N]} \sum_{j:[d]\to[N]} a_{i(1)} \dots a_{i(d)} b_{j(1)} \dots b_{j(d)} \sum_{\substack{\sigma \in \mathbf{S}_N(d) \\ \sigma \in \operatorname{Aut}(i)}} \sum_{\substack{\rho \in \mathbf{S}_N(d) \\ \rho \in \operatorname{Aut}(j)}} (\Gamma^{-1})_{\sigma\rho} \\ &= \sum_{d=0}^{\infty} \frac{z^d}{d!} N^d \sum_{\sigma \in \mathbf{S}_N(d)} \sum_{\rho \in \mathbf{S}_N(d)} (\Gamma^{-1})_{\sigma\rho} \sum_{\substack{i:[d]\to[N] \\ i \in \operatorname{Fix}(\sigma)}} \sum_{\substack{j:[d]\to[N] \\ j \in \operatorname{Fix}(\rho)}} a_{i(1)} \dots a_{i(d)} b_{j(1)} \dots b_{j(d)} \\ &= \sum_{d=0}^{\infty} \frac{z^d}{d!} N^d \sum_{\sigma \in \mathbf{S}_N(d)} \sum_{\rho \in \mathbf{S}_N(d)} (\Gamma^{-1})_{\sigma\rho} p_\sigma(A) p_\rho(B), \end{aligned}$$

where $p_\sigma(A), p_\rho(B)$ denote the power-sum symmetric functions labelled by the cycle types of σ and ρ specialized at the eigenvalues of A and B.

□

1.3. **Resistance matrices.** The propagator in the $\mathbf{U}(N)$-Wick Lemma, i.e. the Gram matrix associated to the orthogonal projection of V_{dd} onto its subspace of $\mathbf{U}(N)$-invariants, may be interpreted as a specialization of the *resistance matrix* associated to a certain metric on the symmetric group $\mathbf{S}(d)$.

Let T be a set of transpositions which generates $\mathbf{S}(d)$, and let $|\pi| = |\pi|_T$ be the corresponding word norm on $\mathbf{S}(d)$. Thus $|\sigma\rho^{-1}|$ is the length of a geodesic joining ρ to σ in the Cayley graph of $\mathbf{S}(d)$ corresponding to the generating set T. This defines a metric on $\mathbf{S}(d)$.

Let q be a complex variable. The resistance matrix $\Omega = \Omega_T(q)$ associated to the T-induced metric on $\mathbf{S}(d)$ is by definition the $d! \times d!$ matrix

$$\Omega = \begin{bmatrix} & \vdots & \\ \dots & q^{|\sigma\rho^{-1}|} & \dots \\ & \vdots & \end{bmatrix}_{\sigma,\rho\in\mathbf{S}(d)}. \tag{1.11}$$

If we think of the Cayley graph as an electrical network in which each edge is a wire of resistivity q, then the (σ,ρ)-entry of Ω represents the compound resistance encountered by a charge traversing any geodesic from ρ to σ.

One may inquire after various properties of the resistance matrix associated to a given metric on $\mathbf{S}(d)$, such as its determinant as a function of q, or the form of the inverse matrix Ω^{-1} away from the zeros of the determinant. Questions of this sort were posed by Zagier [65] in the case that $T = \{(1\ 2), (2\ 3), \dots, (d-1\ d)\}$ are the Coxeter generators of the symmetric group.

For the Coxeter metric, $|\pi|$ is equal to the number of inversions in π. Zagier used this fact to reduce the problem of proving the existence of a Fock space representation for the q-Heisenberg algebra when $-1 < q < 1$ to the problem of proving that the resistance matrix Ω is non-singular in this range. He then proved the invertibility of Ω away from certain roots of unity by observing that it is the image of the *resistance element*

$$\omega = \sum_{\pi\in\mathbf{S}(d)} q^{|\pi|}\pi \tag{1.12}$$

in the regular representation of the group algebra $\mathbb{C}[\mathbf{S}(d)]$, with respect to the standard (permutation) basis. This provides leverage on the problem since, as shown by Zagier, the resistance element can be factored in $\mathbb{C}[\mathbf{S}(d)]$ as

$$\omega = \zeta_1\zeta_2\dots\zeta_d, \tag{1.13}$$

where the factors ζ_i are certain group algebra elements whose images in the irreducible representations V^λ of $\mathbb{C}[\mathbf{S}(d)]$ are easy to understand. Knowledge of the action of the individual factors ζ_i in irreducible representations can then be assembled into knowledge of the action of ω in the regular representation using the isotypic decomposition

$$\mathbb{C}[\mathbf{S}(d)] = \bigoplus_{\lambda\vdash d}(\dim\lambda)V^\lambda \tag{1.14}$$

of the group algebra. Indeed, this is precisely the strategy invented by Frobenius in his solution of Dedekind's group determinant question, which was the initial motivation behind the character theory of finite groups, see [40] as well as the discussion in [65]. In this way Zagier obtained the beautiful formula

$$\det \Omega = \prod_{i=1}^{d-1} (1 - q^{i(i+1)})^{e_i}, \quad e_i = \binom{d}{i+1}(i-1)!(d-i)!, \tag{1.15}$$

from which it is clear that Ω is non-singular provided q is not an $i(i+1)$-st root of unity for any $i = 1, \dots, d-1$.

Later, it was pointed out by Hanlon and Stanley [29] that the resistance matrix Ω may be viewed as a specialization of the Varchenko matrix associated to the A_{d-1} hyperplane arrangement, and so the determinant evaluation (1.15) also follows from general formulas due to Varchenko [63].

In our setting, we are interested in the resistance matrix corresponding to the generating set $T = C_{(21^{d-2})}$, the entire conjugacy class of transpositions. The corresponding word norm is given by $|\pi| = d - c(\pi)$, with $c(\pi)$ the number of cycles in π (including 1-cycles). Thus, for $d \leq N$, the propagator in the $\mathbf{U}(N)$-Wick Lemma is given by

$$\Gamma = N^d \Omega, \tag{1.16}$$

with resistivity specialized at $q = 1/N$.

In order to understand the propagator in the $\mathbf{U}(N)$-Wick formula, we may follow Zagier's strategy and factor the corresponding resistance element $\omega \in \mathbb{C}[\mathbf{S}(d)]$ into tractable pieces. The factorization implementing this strategy has been known since the work of Jucys [36], see [14, Proposition 2.1] for a proof.

Lemma 1.6. *For the all-transpositions distance on* $\mathbf{S}(d)$*, the resistance element* ω *factors as*

$$\omega = (1 + qJ_1)(1 + qJ_2) \dots (1 + qJ_d),$$

where $J_1 := 0$ *and*

$$J_t = \sum_{s<t} (s\ t)$$

for $2 \leq t \leq d$.

The transposition sums appearing in the above lemma are the *Jucys-Murphy elements*:

$$\begin{aligned} J_2 &= (1\ 2) \\ J_3 &= (1\ 3) + (2\ 3) \\ J_4 &= (1\ 4) + (2\ 4) + (3\ 4) \\ &\vdots \end{aligned} \tag{1.17}$$

They can alternatively be written as

$$J_t = \sum \text{transpositions in } \mathbf{S}(t) - \sum \text{transpositions in } \mathbf{S}(t-1), \tag{1.18}$$

which shows that J_t belongs to the (maximal) commutative subalgebra of $\mathbb{C}[\mathbf{S}(d)]$ simultaneously generated by the images of the class algebras $\mathcal{Z}(1), \mathcal{Z}(2), \dots, \mathcal{Z}(d)$ under the standard embedding $\mathbf{S}(k) \hookrightarrow \mathbf{S}(d)$ for $k \leq d$ (this is the Gelfand-Zetlin algbera; see [51]). Note also that the right hand side of the factorization in Lemma 1.6 is precisely the generating function for the elementary symmetric functions,

$$\sum_{r=0}^{\infty} q^r e_k(x_1, x_2, \dots) = \prod_{i=1}^{\infty} (1 + qx_i), \tag{1.19}$$

specialized on the commutative alphabet $\Xi_d = \{\{J_1, J_2, \dots, J_d, 0, 0, \dots\}\}$. Thus an equivalent statement of Lemma 1.6 is

$$e_k(\Xi_d) = \sum_{\substack{\mu \vdash d \\ \ell(\mu) = d-k}} C_\mu, \tag{1.20}$$

where $C_\mu \in \mathcal{Z}(d)$ is the sum of all permutations of cycle type μ. It follows from the fundamental theorem in symmetric function theory, $\Lambda = \mathbb{C}[e_1, e_2, \dots]$, that $f \mapsto f(\Xi_d)$ defines a specialization $\Lambda \to \mathcal{Z}(d)$ mapping the algebra of symmetric functions onto the class algebra.

Since $f(\Xi_d) \in \mathcal{Z}(d)$ for any symmetric function f, it follows from Schur's Lemma that $f(\Xi_d)$ acts as a scalar operator in any irreducible representation V^λ of $\mathbb{C}[\mathbf{S}(d)]$. It is a remarkable result of Jucys [36] that the central character of $f(\Xi_d)$ acting in V^λ is given by the substitution rule $f(\Xi_d) \mapsto f(A_\lambda)$, where

$$A_\lambda = \{\{c(\square) : \square \in \lambda\}\} \tag{1.21}$$

is the alphabet of contents[2] of the Young diagram $\lambda \vdash d$. In particular, the central character of the resistance element ω in V^λ is

$$\prod_{\square \in \lambda} (1 + qc(\square)). \tag{1.22}$$

We now see from the isotypic decomposition $\mathbb{C}[\mathbf{S}(d)] = \bigoplus_{\lambda \vdash d} (\dim \lambda) V^\lambda$ that the determinant of the resistance matrix Ω corresponding to the all-transpositions word metric on $\mathbf{S}(d)$ is

$$\det \Omega = \prod_{c=1}^{d-1} (1 - c^2 q^2)^{e_c}, \quad e_c = \sum_{\substack{\lambda \vdash d \\ c \in A_\lambda}} \dim \lambda. \tag{1.23}$$

It follows from the above computation that the resistance matrix Ω associated to the all-transpositions distance on $\mathbf{S}(d)$ is invertible for $q \notin \{\pm 1, \pm \frac{1}{2}, \dots, \pm \frac{1}{d-1}\}$. The entries of the inverse matrix Ω^{-1} are rational functions of q which are analytic in the disc $|q| < \frac{1}{d-1}$. In particular, by virtue of the reciprocal relationship between the elementary symmetric functions e_r and the complete symmetric functions

[2] Recall that the content of the cell $\square$ in row i and column j of a Young diagram λ is $c(\square) = j - i$.

$$h_r(x_1, x_2, \dots) = \sum_{t_1 \le t_2 \le \cdots \le t_r} x_{t_1} x_{t_2} \dots x_{t_r}, \tag{1.24}$$

the Maclaurin series of the matrix elements of Ω^{-1} is

$$(\Omega^{-1})_{\sigma\rho} = \sum_{r=0}^{\infty} (-q)^r [\sigma\rho^{-1}] h_r(\Xi_d) = (-1)^{|\sigma\rho^{-1}|} \sum_{r=0}^{\infty} q^r [\sigma\rho^{-1}] h_r(\Xi_d), \tag{1.25}$$

with radius of convergence at least $\frac{1}{d-1}$ for each matrix element. Note that the above is *not* an alternating series; its coefficients are either non-negative integers or non-positive integers, depending on the parity of the permutation $\sigma\rho^{-1}$.

If we interpret Ω^{-1} as the "propagator" in the $\mathbf{U}(N)$-Wick formula, we see a major departure from the Gaussian Wick Lemma: the propagator makes power series contributions. This feature of polynomial integrals over $\mathbf{U}(N)$ was first pointed out by De Wit and 't Hooft [13]. The poles of the matrix elements of Ω^{-1} are known as "De Wit - 't Hooft anomalies" in the physics literature [55].

1.4. **Asymptotic expansion of Taylor coefficients.** We now substitute the power series form of the propagator $\Omega^{-1} = N^d \Gamma^{-1}$ into the series expansion of the HCIZ integral obtained in Proposition 1.5. This yields:

$$\begin{aligned} I_N(z) &= \sum_{d=0}^{N} \left(\sum_{\sigma \in \mathbf{S}(d)} \sum_{\rho \in \mathbf{S}(d)} (\Omega^{-1})_{\sigma\rho} p_\sigma(A) p_\rho(B) \right) \frac{z^d}{d!} + \text{higher terms in } z \\ &= \sum_{d=0}^{N} \left(\sum_{\sigma \in \mathbf{S}(d)} \sum_{\rho \in \mathbf{S}(d)} \sum_{r=0}^{\infty} \left(-\frac{1}{N} \right)^r [\sigma\rho^{-1}] h_r(\Xi_d) p_\sigma(A) p_\rho(B) \right) \frac{z^d}{d!} + \text{higher terms in } z \\ &= \sum_{d=0}^{N} \left(\sum_{\alpha,\beta \vdash d} \sum_{r=0}^{\infty} \left(-\frac{1}{N} \right)^r [C_{(1^d)}] C_\alpha C_\beta h_r(\Xi_d) p_\alpha(A) p_\beta(B) \right) \frac{z^d}{d!} + \text{higher terms in } z. \end{aligned}$$

We have arrived at the identity

$$I_N(z) = \sum_{d=0}^{N} \left(\sum_{\alpha,\beta \vdash d} \sum_{r=0}^{\infty} \left(-\frac{1}{N} \right)^r [C_{(1^d)}] C_\alpha C_\beta h_r(\Xi_d) p_\alpha(A) p_\beta(B) \right) \frac{z^d}{d!} + \text{higher terms in } z,$$

which gives an absolutely convergent series expansion for each of the first N coefficients in the Maclaurin series of $I_N(z)$ in terms of multiplication in the center of the symmetric group algebra.

From the definition of the JM elements and the complete symmetric function h_r, one sees that the expression $[C_{(1^d)}] C_\alpha C_\beta h_r(\Xi_d)$ counts $(r+2)$-tuples $(\sigma, \rho, \tau_1, \dots, \tau_r)$ of permutations from the symmetric group $\mathbf{S}(d)$ such that

(1) σ has cycle type α, ρ has cycle type β, and $\tau_1, \dots, \tau_r$ are transpositions;

(2) The product $\sigma\rho\tau_1 \dots \tau_r$ is the identity permutation;

(3) The transpositions $\tau_1 = (s_1\ t_1), \dots, \tau_r = (s_r\ t_r)$ satisfy $t_1 \le \cdots \le t_r$.

Observe that z is an exponential marker for the size d of the ground set in this combinatorial problem, while $-1/N$ is an ordinary marker for the number r of transposition factors, which must be *ordered* according to the monotonicity constraint $t_1 \leq \cdots \leq t_r$. Since

$$e^{N^2 F_N(z)} = I_N(z), \quad F_N(0) = 0, \tag{1.26}$$

it follows from the general theory of generating functions [19, Chapter 3] that the Maclaurin series of $F_N(z)$ is

$$F_N(z) = N^{-2} \sum_{d=0}^{N} \left(\sum_{\alpha,\beta \vdash d} \sum_{r=0}^{\infty} \left(-\frac{1}{N}\right)^r \vec{H}^r(\alpha,\beta) p_\alpha(A) p_\beta(B) \right) \frac{z^d}{d!} + \text{higher terms in } z.$$

where $\vec{H}^r(\alpha,\beta)$ is the number of solutions to the same combinatorial problem as above, but now with the additional condition that the subgroup of $\mathbf{S}(d)$ generated by the factors $\sigma, \rho, \tau_1, \ldots, \tau_r$ acts transitively on the points $\{1, \ldots, d\}$. This is exactly like the passage from all maps to connected maps via the logarithm in Hermitian matrix models. By the Riemann-Hurwitz formula, $\vec{H}^r(\alpha,\beta) = \vec{H}_g(\alpha,\beta)$ when $r = 2g - 2 + \ell(\alpha) + \ell(\beta)$ for $g \geq 0$, and vanishes otherwise. We thus obtain the following convergent series representation for each of the first N derivatives of $F_N(z)$ at $z = 0$.

Theorem 1.7. *For $1 \leq d \leq N$, we have the absolutely convergent series representation*

$$F_N^{(d)}(0) = \sum_{g=0}^{\infty} \frac{C_{g,d,N}}{N^{2g}},$$

where

$$C_{g,d,N} = \sum_{\alpha,\beta \vdash d} (-1)^{\ell(\alpha)+\ell(\beta)} \vec{H}_g(\alpha,\beta) (N^{-\ell(\alpha)} p_\alpha(A)) (N^{-\ell(\beta)} p_\beta(B)).$$

The $N \to \infty$ asymptotic expansion of $F_N^{(d)}(0)$ under the hypotheses of Theorem 0.1 now follows from Theorem 1.7. Let $(A_N), (B_N)$ be two sequences of $N \times N$ normal matrices whose spectral radii are uniformly bounded, with least upper bound M, and which admit limit moments

$$\begin{aligned} -\phi_k &:= \lim_{N \to \infty} N^{-1} \operatorname{tr}(A_N^k) \\ -\psi_k &:= \lim_{N \to \infty} N^{-1} \operatorname{tr}(B_N^k) \end{aligned} \tag{1.27}$$

of all orders. The first of these assumptions implies the inequalities

$$\begin{aligned} |p_\alpha(A_N)| &\leq N^{\ell(\alpha)} M^{|\alpha|} \\ |p_\beta(B_N)| &\leq N^{\ell(\beta)} M^{|\beta|}, \end{aligned} \tag{1.28}$$

while the second is equivalent to the existence of the limits

$$(1.29)\qquad \begin{aligned}(-1)^{\ell(\alpha)}\phi_\alpha &= \lim_{N\to\infty} N^{-\ell(\alpha)} p_\alpha(A_N)\\ (-1)^{\ell(\beta)}\psi_\beta &= \lim_{N\to\infty} N^{-\ell(\beta)} p_\beta(B_N)\end{aligned}$$

where

$$(1.30)\qquad \begin{aligned}\phi_\alpha &= \prod_{i=1}^{\ell(\alpha)} \phi_{\alpha_i}\\ \psi_\beta &= \prod_{j=1}^{\ell(\beta)} \psi_{\beta_j}.\end{aligned}$$

Theorem 1.8. *With the assumptions and notation of Theorem 0.1, we have the $N \to \infty$ asymptotic expansion*

$$F_N^{(d)}(0) \sim \sum_{g=0}^{\infty} \frac{C_{g,d}}{N^{2g}}$$

for the derivatives of $F_N(z)$ at $z = 0$.

Proof. It is required to prove that

$$\lim_{N\to\infty} N^{2h}\left(F_N^{(d)}(0) - \sum_{g=0}^{h-1} \frac{C_{g,d}}{N^{2g}}\right) = C_{h,d}$$

for each $h \geq 0$.

Let $h \geq 0$ be given. By Theorem 1.7, we have

$$N^{2h}\left(F_N^{(d)}(0) - \sum_{g=0}^{h-1} \frac{C_{g,d,N}}{N^{2g}}\right) = C_{h,d,N} + \frac{C_{h+1,d,N}}{N^2} + \dots$$

for all N sufficiently large. Now

$$\begin{aligned}|C_{g,d,N}| &= \left|\sum_{\alpha,\beta\vdash d} (-1)^{\ell(\alpha)+\ell(\beta)} \vec{H}_g(\alpha,\beta)(N^{-\ell(\alpha)}p_\alpha(A_N))(N^{-\ell(\beta)}p_\beta(B_N))\right|\\ &\leq M^{2d}\sum_{\alpha,\beta} \vec{H}_g(\alpha,\beta)\\ &\leq M^{2d} p(d)^2 (d!)^{2g-2+2d},\end{aligned}$$

where $p(d)$ denotes the number of partitions of d and we have used the estimate

$$\vec{H}_g(\alpha,\beta) \leq (d!)^{2g-2+2d},$$

which follows immediately from the definition of the monotone double Hurwitz numbers. Thus

$$\frac{C_{h+1,d,N}}{N^2} + \frac{C_{h+2,d,N}}{N^4} + \dots \leq \frac{M^{2d}p(d)^2(d!)^{2(d+h)}}{N^2}\left(1 + \frac{(d!)^2}{N^2} + \frac{(d!)^4}{N^4} + \dots\right)$$
$$= \frac{M^{2d}p(d)^2(d!)^{2(d+h)}}{N^2}\frac{1}{1-\frac{(d!)^2}{N^2}}$$

for all N sufficiently large, so that

$$N^{2h}\left(F_N^{(d)}(0) - \sum_{g=0}^{h-1}\frac{C_{g,d,N}}{N^{2g}}\right) = C_{h,d,N} + O\left(\frac{1}{N^2}\right)$$

as $N \to \infty$. The result now follows since

$$\lim_{N\to\infty} C_{g,d,N} = C_{g,d}. \tag{1.31}$$

□

1.5. **Convergence of Hurwitz generating functions.** Working formally with Theorem 1.8, which gives the full $N \to \infty$ asymptotic expansion of each Taylor coefficient of $F_N(z)$ about $z = 0$, we are led to the following $N \to \infty$ expansion of $F_N(z)$ itself:

$$F_N(z) = \sum_{d=1}^{\infty} F_N^{(d)}(0)\frac{z^d}{d!} \sim \sum_{d=1}^{\infty}\left(\sum_{g=0}^{\infty}\frac{C_{g,d}}{N^{2g}}\right)\frac{z^d}{d!} \sim \sum_{g=0}^{\infty}\frac{1}{N^{2g}}\left(\sum_{d=1}^{\infty}C_{g,d}\frac{z^d}{d!}\right). \tag{1.32}$$

These formal manipulations suggest the asymptotic expansion of $F_N(z)$ claimed in Theorem 0.1. In order to verify this rigorously, we must determine the radius of convergence of the power series

$$C_g(z) = \sum_{d=1}^{\infty} C_{g,d}\frac{z^d}{d!}. \tag{1.33}$$

If we are able to prove that these series all converge in a common neighbourhood of $z = 0$, Theorem 0.1 will follow.

By the boundedness of spectral radii hypothesis, we have the estimate

$$|C_{g,d}| = \left|\sum_{\alpha,\beta\vdash d}\vec{H}_g(\alpha,\beta)\phi_\alpha\psi_\beta\right| \leq M^{2d}\sum_{\alpha,\beta\vdash d}\vec{H}_g(\alpha,\beta), \tag{1.34}$$

so that our problem reduces to determining the radius of convergence of the fixed-genus generating functions

$$\vec{\mathbf{H}}_g(z) = \sum_{d=1}^{\infty}\left(\sum_{\alpha,\beta\vdash d}\vec{H}_g(\alpha,\beta)\right)\frac{z^d}{d!} \tag{1.35}$$

of the monotone double Hurwitz numbers. The convergence of fixed-genus generating functions of Hurwitz numbers is an interesting problem which seems not to have been addressed in the literature on Hurwitz theory. We will solve it by reducing to the case of simple Hurwitz numbers, and then applying a specialization of the results in our second paper [18].

Definition 1.9. The *monotone simple Hurwitz number* $\vec{H}_{g,d} := \vec{H}_g((1^d),(1^d))$ is equal to the number of r-tuples $(\tau_1,\ldots,\tau_r)$ of transpositions from the symmetric group $\mathbf{S}(d)$ such that:

1. The product $\tau_1 \ldots \tau_r$ equals the identity permutation;
2. The group $\langle \tau_1,\ldots,\tau_r \rangle$ acts transitively on $\{1,\ldots,d\}$;
3. $r = 2g-2+2d$;
4. Writing $\tau_i = (s_i \; t_i)$ with $s_i < t_i$, we have $t_1 \leq \cdots \leq t_r$.

The monotone simple Hurwitz number $\vec{H}_{g,d}$ counts, up to isomorphism, a combinatorially restricted subclass of the set of genus g, degree d branched covers of $\mathbb{P}^1$ with simple ramification over $r = 2g-2+2d$ fixed points of the sphere and no other branching, where the number r is determined by the Riemann-Hurwitz formula. The fixed-genus generating function encoding the monotone simple Hurwitz numbers in genus g is

$$\vec{\mathbf{H}}^s_g(z) = \sum_{d=1}^{\infty} \vec{H}_{g,d} \frac{z^d}{d!}. \tag{1.36}$$

Theorem 1.10. *The generating functions $\vec{\mathbf{H}}_g(z)$ and $\vec{\mathbf{H}}^s_g(z)$ have the same radius of convergence.*

Proof. Clearly, we have that

$$\sum_{\alpha,\beta \vdash d} \vec{H}_g(\alpha,\beta) \geq \vec{H}_{g,d},$$

so the radius of convergence of $\vec{\mathbf{H}}_g(z)$ is at most the radius of convergence of $\vec{\mathbf{H}}^s_g(z)$.

Conversely, a straightforward combinatorial argument involving successive multiplication by appropriately ordered cut transpositions [18] leads to the inequality

$$\vec{H}_g(\alpha,\beta) \leq \vec{H}_g(21^{d-2},21^{d-2}) \text{ for all } \alpha,\beta \vdash d$$

as well as the identity

$$\vec{H}_g(21^{d-2},21^{d-2}) = \frac{d^2}{4}\vec{H}_{g,d},$$

so that we have

$$\sum_{\alpha,\beta \vdash d} \vec{H}_g(\alpha,\beta) \leq \frac{d^2 p(d)^2}{4}\vec{H}_{g,d}$$

where $p(d)$ is the number of partitions of d. By the Hardy-Ramanujan formula,

$$p(d) \sim \frac{e^{\pi\sqrt{\frac{2}{3}}\sqrt{d}}}{4\sqrt{3}d}, \quad d \to \infty,$$

we have

$$\lim_{d\to\infty}\left(\frac{d^2p(d)^2}{4}\right)^{1/d}=1.$$

Thus the radius of convergence of $\vec{\mathbf{H}}_g(z)$ is at least the radius of convergence of $\vec{\mathbf{H}}^s_g(z)$.

□

We can now combine Theorem 1.10 with the results of our second paper [18] to determine the radius of convergence of the generating functions $\vec{\mathbf{H}}_g(z)$.

In genus $g=0$, this is a very direct and tangible calculation. Specializing $\alpha=(1^d)$ in [18, Theorem 0.3] yields the exact formula

$$\frac{\vec{H}_{0,d}}{d!}=\frac{2^{d-1}}{d^2(2d-1)}\binom{3d-3}{d-1} \tag{1.37}$$

for the genus zero monotone simple Hurwitz numbers, from which we directly find that the radius of convergence of $\vec{\mathbf{H}}^s_0(z)$, and hence $\vec{\mathbf{H}}_0(z)$, is 2/27. Note that this value was also determined by Zinn-Justin [66] using the dispersionless Toda formalism.

In genus $g\geq 1$, specializing [18, Theorem 0.5] at $\alpha=(1^d)$ yields the following rational form.

Theorem 1.11. *Let*

$$s(z)=\sum_{n=1}^{\infty}\frac{2^{n-1}}{n}\binom{3n-2}{n-1}z^n$$

be the unique formal power series solution of the functional equation

$$s=z(1-2s)^{-2}$$

in the ring $\mathbb{C}[[z]]$, *obtained by Lagrange Inversion.*

For $g=1$, *we have*

$$\vec{\mathbf{H}}^s_1(z)=\frac{1}{8}\log(1-2s)-\frac{1}{24}\log(1-6s),$$

while for any $g\geq 2$, *we have*

$$\vec{\mathbf{H}}^s_g(z)=-c_{g,(0)}+\frac{1}{(1-6s)^2}\sum_{d=0}^{3}\sum_{\alpha\vdash d}\frac{c_{g,\alpha}6^{\ell(\alpha)}s^{\ell(\alpha)}}{(1-6s)^{\ell(\alpha)}},$$

where $c_{g,\alpha}\in\mathbb{Q}$ *are rational constants.*

It is clear from Theorem 1.11 that each generating function $\vec{\mathbf{H}}^s_g(z), g\geq 1$, has radius of convergence 2/27.

Remark 1.12. Considering in more detail the genus zero case, Zinn-Justin [66, p. 425] observes that by comparing with results in Tutte [62], one finds

$$z\frac{\partial}{\partial z}\vec{\mathbf{H}}^s_0(z)=z+zp(z)=\sum_{d=1}^{\infty}\frac{2^{d-1}}{d^2(2d-1)}\binom{3d-3}{d-1}z^d, \tag{1.38}$$

where $p(z)$ is the generating series for rooted planar maps with respect to vertices. Tutte specifies $p(z)$ parametrically by

$$p = \phi(1-2\phi), \qquad \phi = z(1+2\phi)^3. \tag{1.39}$$

The results in our second paper [18, Section 3] specify $\vec{\mathbf{H}}_0^s(z)$ parametrically by

$$\left(2z\frac{\partial}{\partial z} - 1\right) z\frac{\partial}{\partial z}\vec{\mathbf{H}}_0^s(z) = s - s^2, \qquad s = z(1-2s)^{-2}. \tag{1.40}$$

These parameterizations can be reconciled by the transformation

$$s = \frac{\phi}{1+2\phi}. \tag{1.41}$$

1.6. **Asymptotic expansion of the free energy.** We now know that the generating functions

$$C_g(z) = \sum_{d=1}^{\infty} C_{g,d}\frac{z^d}{d!}, \tag{1.42}$$

which encode the orders of the $N \to \infty$ asymptotic expansion of the Taylor coefficients of $F_N(z)$ about $z = 0$, all converge absolutely in the open disc $D(0, r_c M^{-2})$, where r_c is the critical value $2/27$. The proof of Theorem 0.1 follows from this by taking

$$G_N(z) = N^{2h}\left(F_N(z) - \sum_{g=0}^{h-1}\frac{C_g(z)}{N^{2h}}\right) \tag{1.43}$$

in the following proposition.

Proposition 1.13. *Let (G_N) be a sequence of complex functions, each holomorphic in a neighbourhood of $z = 0$. Suppose that the radius of convergence of the Maclaurin series*

$$G_N(z) = \sum_{d=0}^{\infty} G_N^{(d)}(0)\frac{z^d}{d!}$$

of G_N is $r_N > 0$. Suppose also that the limits

$$L_d := \lim_{N\to\infty} G_N^{(d)}(0), \quad d \geq 0,$$

exist and that the power series

$$L(z) = \sum_{d=0}^{\infty} L_d\frac{z^d}{d!}$$

has radius of convergence $r > 0$. Then

(1) $\lim_{N\to\infty} r_n = r$.
(2) *G_N converges to L pointwise on $D(0, r)$.*
(3) *G_N converges to L uniformly on compact subsets of $D(0, r)$.*

Proof. (1) Follows from the Cauchy-Hadamard theorem.

(2) Let z_0 be an arbitrary point in $D(0,r)$. By (1), $z_0 \in D(0,r_N)$ for all N sufficiently large. Hence, for N sufficiently large,

$$\sum_{d=0}^{\infty}(G_N^{(d)}(0) - L_d)\frac{z_0^d}{d!}$$

is an absolutely convergent series of complex numbers whose sum is $G_N(z_0) - L(z_0)$. Let $\varepsilon > 0$ be given. For any non-negative integer E, we have the inequality

$$\left|\sum_{d=0}^{\infty}(G_N^{(d)}(0) - L_d)\frac{z_0^d}{d!}\right| \leq \sum_{d=0}^{E}\left|G_N^{(d)}(0) - L_d\right|\frac{|z_0|^d}{d!} + \left|\sum_{d=E+1}^{\infty}(G_N^{(d)}(0) - L_d)\frac{z_0^d}{d!}\right|.$$

The second group of terms is the tail of a convergent series, so we may choose E sufficiently large (depending on z_0 and ε) so that it is $< \varepsilon/2$. The first group of terms is a finite sum each of whose terms go to zero as $N \to \infty$, so we may choose N sufficiently large (depending on z_0, ε, and E) so that it is $< \varepsilon/2$. Thus $|G_N(z_0) - L(z_0)| < \varepsilon$ for all N sufficiently large, as required.

(3) Let $K \subset D(0,r)$ be a compact set. By (1), $K \subset D(0,r_N)$ for all N sufficiently large. Thus $G_N(z) - L(z)$ is holomorphic on K for N sufficiently large, and the result follows from (2) and the Extreme Value Theorem.

□

2. Structure of Monotone Double Hurwitz Numbers

Having established that the monotone double Hurwitz numbers are the combinatorial/geometric objects underlying the notion of genus expansion in the HCIZ model, it is of interest to gather as much information as possible regarding the structure of these objects. In particular, it is natural to seek analogues of the structural properties of the usual double Hurwitz numbers in this new setting. We are thus motivated to search for integrable properties of the Witten-type generating function for monotone double Hurwitz numbers in all degrees and genera, and to look for polynomial behaviour exhibited by the monotone double Hurwitz numbers themselves, regarded as functions on pairs of partitions.

2.1. **Integrable hierarchies.**

2.1.1. The link between Hurwitz theory and integrable systems was suggested by Pandharipande [53], who showed that the (at the time conjectural) Toda equation for the Gromov-Witten potential of $\mathbb{P}^1$ implies a Toda equation satisfied by the generating function

$$\mathbf{H}^s(z,q) = \sum_{d=1}^{\infty}\frac{z^d}{d!}\sum_{r=0}^{\infty}\frac{q^r}{r!}H_d^r, \tag{2.1}$$

of the classical simple Hurwitz numbers in all degrees and genera. Here $H_d^r = H_{g,d}$ counts branched covers of $\mathbb{P}^1$ by curves of genus g with $r = 2g-2+2d$ simple ramification points at fixed positions. Pandharipande's conjecture was settled in short order by Okounkov [46], who proved the considerably stronger result that the richer generating function

$$\mathbf{H}(z,q,A,B) = \sum_{d=1}^{\infty} \frac{z^d}{d!} \sum_{r=0}^{\infty} \frac{q^r}{r!} \sum_{\alpha,\beta} H^r(\alpha,\beta) p_\alpha(A) p_\beta(B) \tag{2.2}$$

for the double Hurwitz numbers $H^r(\alpha,\beta) = H_g(\alpha,\beta), r = 2g-2+\ell(\alpha)+\ell(\beta)$, in which arbitrary ramification is permitted over 0 and ∞ in addition to the r simple ramification points, satisfies the entire 2D Toda lattice hierarchy of Takasaki and Ueno [60] in the variables $p_1(A), p_2(A), \dots, p_1(B), p_2(B), \dots$, where

$$\begin{array}{ll} p_1(A) = a_1 + a_2 + \dots & p_1(B) = b_1 + b_2 + \dots \\ p_2(A) = a_1^2 + a_2^2 + \dots & p_2(B) = b_1^2 + b_2^2 + \dots \\ \vdots & \vdots \end{array}$$

are the power-sum symmetric functions in auxiliary sets of variables $A = \{a_1, a_2, \dots\}$ and $B = \{b_1, b_2, \dots\}$. This evidence of structure in the double Hurwitz numbers is what led to their further investigation by Goulden, Jackson and Vakil [23] and many other authors since, see the discussion of piecewise polynomiality below.

In this section we prove Theorem 0.2, which is the monotone analogue of Okounkov's result cited above. Introduce the generating function

$$\vec{\tau}(z,q,A,B) = \sum_{d=0}^{\infty} \frac{z^d}{d!} \sum_{r=0}^{\infty} q^r \sum_{\alpha,\beta \vdash d} [C_{(1^d)}] C_\alpha C_\beta h_r(\Xi_d) p_\alpha(A) p_\beta(B), \tag{2.3}$$

which is the partition function counting all (i.e. including possibly disconnected) monotone double Hurwitz covers of $\mathbb{P}^1$. As in Section 1, the coefficient

$$\left[\frac{z^d}{d!} q^r p_\alpha(A) p_\beta(B) \right] \vec{\tau}(z,q,A,B) = [C_{(1^d)}] C_\alpha C_\beta h_r(\Xi_d) \tag{2.4}$$

of this formal power series is equal to the number of $(r+2)$-tuples[3] $(\sigma, \rho, \tau_1, \dots, \tau_r)$ of permutations from the symmetric group $\mathbf{S}(d)$ such that:

(1) σ has cycle type α, ρ has cycle type β, and the τ_i are transpositions;

(2) The product $\sigma\rho\tau_1 \dots \tau_r$ equals the identity permutation;

(3) Writing $\tau_i = (s_i\ t_i)$ with $s_i < t_i$, we have $t_1 \le \dots \le t_r$.

The variable z is an exponential marker for the size d of the ground set $\{1, \dots, d\}$, while q is an *ordinary* marker for the number of transposition factors, which must be *ordered* as in condition (3) above. Note that this differs from the generating function for possibly disconnected classical Hurwitz numbers, which is exponential in both the degree of the covering and the number of simple ramification points. By the Exponential Formula (a.k.a. Moment-Cumulant Formula), we have

$$\log \vec{\tau}(z,q,A,B) = \vec{\mathbf{H}}(z,q,A,B) = \sum_{d=1}^{\infty} \frac{z^d}{d!} \sum_{r=0}^{\infty} q^r \sum_{\alpha,\beta \vdash d} \vec{H}^r(\alpha,\beta) p_\alpha(A) p_\beta(B). \tag{2.5}$$

[3]We hope that the use of the Greek letter τ for both tau function and transposition will cause no confusion.

As in [46], we will give a representation-theoretic proof that $\vec{\tau}(z,q,A,B)$ is a tau function of the 2D Toda lattice hierarchy in the variables $p_1(A), p_2(A), \dots, p_1(B), p_2(B), \dots$. The basic ingredient required in the argument, namely Jucys' result on the spectra of symmetric functions of JM elements in irreducible representations of the symmetric group, was already introduced and utilized in Section 1 where we proved that the asymptotic expansion of the d-th Taylor coefficient $F_N^{(d)}(0)$ of the HCIZ free energy is essentially the ordinary generating function of monotone double Hurwitz numbers in fixed degree d.

Let

$$f_\mu(\lambda) = |C_\mu| \frac{\chi^\lambda_\mu}{\dim \lambda} \tag{2.6}$$

denote the central character of the conjugacy class $C_\mu \in \mathcal{Z}(d)$ acting in V^λ, and recall that the eigenvalue of $h_r(\Xi_d)$ acting in V^λ is $h_r(A_\lambda)$. From the isotypic decomposition of $\mathbb{C}[\mathbf{S}(d)]$ we have

$$\begin{aligned} \frac{1}{d!}[C_{(1^d)}]C_\alpha C_\beta h_r(\Xi_d) &= \sum_{\lambda \vdash d} f_\alpha(\lambda) f_\beta(\lambda) h_r(A_\lambda) \left(\frac{\dim \lambda}{d!}\right)^2 \\ &= \frac{|C_\alpha|}{d!}\frac{|C_\beta|}{d!} \sum_{\lambda \vdash d} \chi^\lambda_\alpha \chi^\lambda_\beta h_r(A_\lambda). \end{aligned} \tag{2.7}$$

Using this fact together with the change of basis from the Schur functions to the Newton power-sums,

$$s_\lambda(A) = \sum_{\mu \vdash d} \frac{C_\mu}{d!} \chi^\lambda_\mu p_\mu(A), \tag{2.8}$$

we see that the generating function $\vec{\tau}$ may be rewritten as

$$\vec{\tau}(z,q,A,B) = \sum_\lambda \left(\prod_{\Box \in \lambda} \frac{z}{1 - qc(\Box)} \right) s_\lambda(A) s_\lambda(B), \tag{2.9}$$

where the summation is over all partitions λ and we have used the generating function

$$\sum_{r=0}^\infty h_r(x_1, x_2, \dots) q^r = \prod_{i=1}^\infty \frac{1}{1 - qx_i} \tag{2.10}$$

for the complete homogeneous symmetric functions.

The formula (2.9) already suffices to identify the formal power series $\vec{\tau}(z,q,A,B)$ as a tau function of the KP hierarchy. In particular, it is known that a formal power series of the form

$$\tau = \sum_\lambda Y_\lambda s_\lambda(A) \tag{2.11}$$

is a tau function of the KP hierarchy in the variables $p_1(A), p_2(A), \dots$ if and only if the coefficients Y_λ satisfy the Plücker relations, see e.g. [12]. The coefficients Y_λ

are called the Plücker coordinates of the corresponding tau function. The Schur functions themselves satisfy the Plücker relations and, as shown in [21] so do the products

$$Y_\lambda = \left(\prod_{\Box \in \lambda} y_{c(\Box)} \right) s_\lambda(B), \tag{2.12}$$

where $\{y_0, y_{\pm 1}, y_{\pm 2}, \dots\}$ is an auxiliary set of variables. According to (2.9), the generating function $\vec{\tau}(z, q, A, B)$ is precisely of this form with

$$y_{c(\Box)} = \frac{z}{1 - qc(\Box)}. \tag{2.13}$$

Thus $\vec{\mathbf{H}}(z, q, A, B) = \log \vec{\tau}(z, q, A, B)$ is a solution of the KP hierarchy in the variables $p_1(A), p_2(A), \dots$.

We now wish to prove the stronger statement claimed in Theorem 0.2, namely that $\vec{\mathbf{H}}(z, q, A, B)$ is a solution of the 2D Toda lattice equations. Let us recall that a formal power series solution of the 2D Toda hierarchy in the two sets of variables $p_1(A), p_2(A), \dots, p_1(B), p_2(B), \dots$ of the form

$$\log \left(\sum_\lambda Y_\lambda s_\lambda(A) s_\lambda(B) \right) \tag{2.14}$$

is referred to as a *diagonal solution.* The following recent result due to Carrell [7] asserts that diagonal series whose coefficients Y_λ are shifted content products are solutions of the 2D Toda hierarchy.

Theorem 2.1 ([7])**.** *The sequence of formal power series*

$$\tau_n(z, q, A, B) = \sum_\lambda Y_\lambda(n) s_\lambda(A) s_\lambda(B), \quad n \in \mathbb{Z},$$

where

$$Y_n(\lambda) = \theta_n \prod_{\Box \in \lambda} y_{n + c(\Box)}, \quad n \in \mathbb{Z},$$

and

$$\theta_n = \begin{cases} y_0^{n/2} \prod_{i=1}^{n-1} y_i^{n-i}, \ n > 0 \\ 1, \ n = 0 \\ y_0^{-n/2} \prod_{i=1}^{n-1} y_{-i}^{-n-i}, \ n < 0 \end{cases},$$

is a sequence of tau functions of the 2D Toda lattice equations in the variables $p_1(A), p_2(A), \dots, p_1(B), p_2(B), \dots$.

In light of Theorem 2.1, the same argument we used to demonstrate that $\vec{\mathbf{H}}(z, q, A, B)$ is a solution of the KP hierarchy in the variables $p_1(A), p_2(A), \dots$ also implies that $\vec{\mathbf{H}}(z, q, A, B)$ is the $n = 0$ term of a sequence of diagonal solutions of the 2D Toda lattice equations in $p_1(A), p_2(A), \dots, p_1(B), p_2(B), \dots$.

2.2. **Piecewise polynomiality.** The remarkable ELSV formula [15], which expresses the single Hurwitz numbers $H_g(\alpha)$ as integrals over the Deligne-Mumford compactification $\overline{\mathcal{M}}_{g,m}$ of the moduli space of genus g curves with m marked points, implies the existence of a family of polynomials $P_g(x_1, \ldots, x_m)$ such that

$$\frac{|\operatorname{Aut}\alpha|}{|\alpha|!} H_g(\alpha) = C(g,\alpha) P_g(\alpha_1, \ldots, \alpha_m) \tag{2.15}$$

for all partitions α with $\ell(\alpha) = m$, where the combinatorial prefactor $C(g, \alpha)$ is given explicitly by

$$C(g,\alpha) = (2g-2+m+d)! \prod_{i=1}^{m} \frac{\alpha_i^{\alpha_i}}{\alpha_i!}. \tag{2.16}$$

Goulden, Jackson and Vakil [22] proved that this polynomiality property of single Hurwitz numbers is equivalent, for genus $g \geq 2$, to the rationality of the generating function

$$\mathbf{H}_g(z, A) = \sum_{d=1}^{\infty} \left(\sum_{\alpha \vdash d} \frac{H_g(\alpha)}{(2g-2+\ell(\alpha)+d)!} p_\alpha(A) \right) \frac{z^d}{d!} \tag{2.17}$$

in terms of an implicitly defined set of variables obtained by Lagrange inversion. This rational form had been conjectured by Goulden and Jackson [20] (and proved in genus $g = 2, 3$) prior to the advent of the ELSV formula, but continues to resist a proof independent of ELSV.

In the sequel to this paper, we prove by direct combinatorial methods [18, Theorem 0.5] that the generating function

$$\vec{\mathbf{H}}_g(z, A) = \sum_{d=1}^{\infty} \left(\sum_{\alpha \vdash d} \vec{H}_g(\alpha) p_\alpha(A) \right) \frac{z^d}{d!} \tag{2.18}$$

for the single monotone Hurwitz numbers in genus $g \geq 2$ is rational in an implicitly defined set of Lagrangian variables. This is equivalent to an ELSV-type polynomiality property for monotone single Hurwitz numbers [18, Theorem 0.6], obtained without recourse to an analogue of the ELSV formula: there exist polynomials $\vec{P}_g(x_1, \ldots, x_m)$ such that

$$\frac{|\operatorname{Aut}\alpha|}{|\alpha|!} \vec{H}_g(\alpha) = \vec{C}(\alpha) \vec{P}_g(\alpha_1, \ldots, \alpha_m), \tag{2.19}$$

for all partitions α with $\ell(\alpha) = m$, where the genus-independent combinatorial prefactor is given explicitly by a product of central binomial coefficients

$$\vec{C}(\alpha) = \prod_{i=1}^{m} \binom{2\alpha_i}{\alpha_i} \tag{2.20}$$

over the parts of α.

Polynomiality does not persist for the double Hurwitz numbers $H_g(\alpha, \beta)$, whose structure is much more complicated than that of the single Hurwitz numbers. However, Goulden, Jackson and Vakil [23] showed there is a suitable replacement for polynomiality in this context: piecewise polynomiality. The flavour of this result is as follows. For fixed m, n we may view pairs of partitions (α, β) with $|\alpha| = |\beta|$ and $\ell(\alpha) = m, \ell(\beta) = n$ as the lattice points of the region

$$\mathfrak{R}_{m,n} = \left\{ (x_1, \dots, x_m, y_1, \dots, y_n) \in \mathbb{R}^{m+n}_{>0} : \sum_{i=1}^{n} x_i = \sum_{j=1}^{m} y_j \right\}, \tag{2.21}$$

and for fixed g we may view the double Hurwitz numbers $H_g(\alpha, \beta)$ as defining a function

$$(\alpha, \beta) \mapsto \frac{|\operatorname{Aut} \alpha| |\operatorname{Aut} \beta|}{|\alpha|!} H_g(\alpha, \beta) \tag{2.22}$$

on this set of lattice points. Goulden, Jackson and Vakil [23, Theorem 2.1] proved that there exists a hyperplane arrangement in $\mathbb{R}^{m+n}$ and a collection of polynomials $P_{g,\mathfrak{c}}(x_1, \dots, x_m, y_1, \dots, y_n)$ indexed by the chambers $\mathfrak{c}$ of this arrangement such that the function (2.22) is given by

$$(\alpha, \beta) \mapsto P_{g,\mathfrak{c}}(\alpha_1, \dots, \alpha_m, \beta_1, \dots, \beta_n) \tag{2.23}$$

for all $(\alpha, \beta) \in \mathfrak{c}$. They used this piecewise polynomiality property to motivate and support a conjectural analogue of the ELSV formula for double Hurwitz numbers [23, Conjecture 3.5].

Following [23], the piecewise polynomial structure of the double Hurwitz numbers was investigated by a number of authors [8, 35, 56, 57]. In particular, Johnson [35] exhibits an elegant approach to piecewise polynomiality based on the formalism of the infinite wedge space [50]. This representation-theoretic approach was recently extended by Shadrin, Spitz and Zvonkine [57], where piecewise polynomiality was linked to the theory of shifted symmetric functions in a very structured way. We will use this viewpoint to elucidate the piecewise polynomial structure of the monotone double Hurwitz numbers.

Let $\mathcal{F}(\mathbb{Y})$ denote the algebra of functions $\mathbb{Y} \to \mathbb{C}$ on Young's lattice, and consider the subalgebra Λ^* of $\mathcal{F}$ freely generated by the functions

$$p_k^*(\lambda) = \sum_{i=1}^{\infty} \left[\left(\lambda_i - i + \frac{1}{2} \right)^k - \left(-i + \frac{1}{2} \right)^k \right], \quad k \geq 1. \tag{2.24}$$

The algebra Λ^* is known as the algebra of shifted symmetric functions, and the generators $p_k^*(\lambda)$ are the shifted power-sum symmetric functions. Shifted symmetric functions $f(\lambda)$ are polynomial functions in the parts $\lambda_1, \lambda_2, \dots$ of the input partition which become symmetric after the change of variables $\tilde{\lambda}_i := \lambda_i - i + \frac{1}{2}$.

There are several ways in which the introduction of the algebra Λ^* may be motivated. One reason for the ubiquity of this algebra in the representation theory of the symmetric groups is the Kerov-Olshanski Theorem [38], which asserts that the central characters

$$f_\mu(\lambda) = \frac{|C_\mu|}{\dim \lambda} \chi^\lambda_\mu \tag{2.25}$$

are a linear basis of Λ^*. Another motivation is the deep analogy between random partitions and random Hermitian matrices [47], in which the shifted power-sums $p_k^*(\lambda)$ play the same role as the moments $\mathrm{tr}(A^k)$ of a matrix. Finally, shifted symmetric functions provide a framework in which the precise relationship between the Hurwitz theory and the Gromov-Witten theory of $\mathbb{P}^1$ may be described [50].

Consider the transform $T : \mathcal{F}(\mathbb{Y}) \to \mathcal{F}(\mathbb{Y} \times \mathbb{Y})$ sending functions on partitions to functions on pairs of partitions defined by

$$T(f)(\alpha, \beta) := \frac{1}{\prod \alpha_i} \frac{1}{\prod \beta_j} \sum_{|\lambda|=|\alpha|} \chi_\alpha^\lambda \chi_\beta^\lambda f(\lambda). \tag{2.26}$$

This definition assumes that $|\alpha| = |\beta|$; if this is not the case, i.e. if $|\alpha| > |\beta|$ or vice versa, complete the smaller partition by adding an appropriate number of 1's. For fixed m, n we may view pairs of partitions (α, β) with $|\alpha| = |\beta|$ and $\ell(\alpha) = m, \ell(\alpha) = n$ as lattice points of the region $\mathfrak{R}_{m,n}$, as above. For each pair of proper subsets $I \subset [m], J \subset [n]$, consider the hyperplane

$$W_{I,J} = \left\{ (x_1, \dots, x_m, y_1, \dots, y_n) \in \mathfrak{R}_{m,n} : \sum_{i \in I} x_i = \sum_{j \in J} y_j \right\}. \tag{2.27}$$

This hyperplane arrangement is called the *resonance arrangement* in [35, 57]. A *chamber* of the resonance arrangement is a connected component $\mathfrak{c}$ of $\mathfrak{R}_{m,n} \setminus \bigcup_{I \subset [m], J \subset [n]} W_{I,J}$. The following general piecewise polynomiality property of the transform (2.26) restricted to the algebra of shifted symmetric functions is proved in [57].

Theorem 2.2. *To each triple (f, m, n) consisting of a shifted symmetric function f and a pair of positive integers m, n, there corresponds a collection of polynomials $P_{f,\mathfrak{c}}(x_1, \dots, x_m, y_1, \dots, y_n)$ indexed by the chambers of the resonance arrangement such that*

$$T(f)(\alpha, \beta) = P_{f,\mathfrak{c}}(\alpha_1, \dots, \alpha_m, \beta_1, \dots, \beta_n)$$

for all $(\alpha, \beta) \in \mathfrak{c}$.

The proof of the piecewise polynomiality of double Hurwitz numbers follows directly from Theorem 2.2. Indeed, a straightforward argument as in [57] verifies that when $(\alpha, \beta) \in \mathfrak{c}$, the transitivity condition in the definition of the double Hurwitz numbers is automatically satisfied. Thus double Hurwitz numbers are given by the character formula

$$\frac{1}{|\alpha|!} H_g(\alpha, \beta) = \frac{|C_\alpha|}{|\alpha|!} \frac{|C_\beta|}{|\beta|!} \sum_{|\lambda|=|\alpha|} \chi_\alpha^\lambda \chi_\beta^\lambda (f_2(\lambda))^r \tag{2.28}$$

for all $(\alpha, \beta) \in \mathfrak{c}$, where $r = 2g - 2 + \ell(\alpha) + \ell(\beta)$ and $f_2(\lambda)$ is the central character of the conjugacy class of transpositions in the irreducible representation of the symmetric group labelled by λ. But this central character is a shifted symmetric function:

$$f_2(\lambda) = \frac{1}{2} p_2^*(\lambda). \tag{2.29}$$

We thus have

$$\frac{|\operatorname{Aut}\alpha||\operatorname{Aut}\beta|}{|\alpha|!}H_g(\alpha,\beta) = T\left(\frac{1}{2}p_2^*\right)^r(\alpha,\beta) \tag{2.30}$$

for all $(\alpha,\beta) \in \mathfrak{c}$, with $r = 2g-2+\ell(\alpha)+\ell(\beta)$, so that piecewise polynomiality follows immediately from Theorem 2.2.

The above argument applies verbatim to the monotone double Hurwitz numbers, which are given by the character formula

$$\frac{1}{|\alpha|!}\vec{H}_g(\alpha,\beta) = \frac{|C_\alpha|}{|\alpha|!}\frac{|C_\beta|}{|\beta|!}\sum_{|\lambda|=|\alpha|}\chi^\lambda_\alpha\chi^\lambda_\beta h_r(A_\lambda) \tag{2.31}$$

for all $(\alpha,\beta) \in \mathfrak{c}$, with $r = 2g-2+\ell(\alpha)+\ell(\beta)$, once we verify that the function $\hat{h}_r(\lambda) := h_r(A_\lambda)$ is shifted symmetric. This fact in turn follows an alternative characterization, due to Kerov, of the algebra Λ^* in terms of the contents rather than the parts of partitions (see [52, Proposition 2.4] for a proof).

Theorem 2.3. *The algebra Λ^* coincides with the algebra of functions on partitions generated by $p_1^*(\lambda) = |\lambda|$ and the functions*

$$\hat{p}_k(\lambda) := p_k(A_\lambda), \quad k \geq 1,$$

where $p_k(A_\lambda)$ denotes the usual power-sum symmetric function $p_k \in \Lambda$ specialized on the content alphabet A_λ of λ.

Indeed, one may check directly that $\frac{1}{2}p_2^*(\lambda) = \hat{p}_1(\lambda) = \hat{h}_1(\lambda)$, so that the character formula for the usual double Hurwitz numbers is given by the T-transform of $\hat{h}_{(1^r)}$ while the character formula of the monotone double Hurwitz numbers is given by the T-transform of $\hat{h}_r$.

References

1. J. Baik, E. M. Rains, *Algebraic aspects of increasing subsequences*, Duke Mathematical Journal **109**(1) (2001), 1-65.
2. T. Banica, B. Collins, *Integration over compact quantum groups*, Publ. Res. Inst. Math. Sci. **43** (2007), 277-302.
3. T. Banica, S. Curran, R. Speicher, *Stochastic aspects of easy quantum groups*, Probability Theory and Related Fields **149** (2011), 435-462.
4. D. Bessis, C. Itzykson, J.-B. Zuber, *Quantum field theory techniques in graphical enumeration*, Advances in Applied Mathematics **1** (1980), 109-157.
5. P. Bleher, A. Deaño, *Topological expansion in the cubic random matrix model*, arXiv:1011.6338
6. P. M. Bleher, A. R. Its, *Asymptotics of the partition function of a random matrix model*, Annales de L'Institut Fourier **55**(6) (2005), 1943-2000.
7. S. R. Carrell, *The Bernstein operator and diagonal solutions of integrable hierarchies*, in preparation.
8. R. Cavalieri, P. Johnson, H. Markwig, *Chamber structure of double Hurwitz numbers*, arXiv:1003.1805
9. B. Collins, *Moments and cumulants of polynomial random variables on unitary groups, the Itzykson-Zuber integral, and free probability*, International Mathematics Research Notices **17** (2003), 954-982.
10. B. Collins, A. Guionnet, E. Maurel-Segala, *Asymptotics of unitary and orthogonal matrix integrals*, Advances in Mathematics **222** (2009), 172-215.

11. B. Collins, P. Śniady, *Integration with respect to the Haar measure on unitary, orthogonal and symplectic group*, Communications in Mathematical Physics **264** (2006), 773-795.
12. E. Date, M. Jimbo, T. Miwa, *Solitons: Differential Equations, Symmetries and Infinite Dimensional Lie Algebras*, Cambridge Tracts in Mathematics **135** (2000). Translated from the Japanese by Miles Reid.
13. B. De Wit, G. 't Hooft, *Non-convergence of the $1/N$ expansion for $\mathbf{SU}(N)$ gauge fields on a lattice*, Physics Letters **69**(1) (1977), 61-64.
14. P. Diaconis, C. Greene, *Applications of Murphy's elements*, Stanford University technical report **335** (1989).
15. T. Ekedahl, S. Lando, M. Shapiro, A. Vainshtein, *Hurwitz numbers and intersections on moduli spaces of curves*, Inventiones Mathematicae **146** (2001), 297-327.
16. N. M. Ercolani, *Caustics, counting maps and semiclassical asymptotics*, Nonlinearity **24** (2011), 481-526.
17. N. M. Ercolani, K. D. T.-R. McLaughlin, *Asymptotics of the partition function for random matrices via Riemann-Hilbert techniques and applications to graphical enumeration*, International Mathematics Research Notices **14** (2003), 755-820.
18. I. P. Goulden, M. Guay-Paquet, J. Novak, *Monotone Hurwitz numbers and the HCIZ integral II*, arxiv preprint (2011).
19. I. P. Goulden, D. M. Jackson, *Combinatorial Enumeration*, John Wiley and Sons, New York, 1983 (reprinted by Dover, 2004).
20. I. P. Goulden, D. M. Jackson, *Numer of ramified covers of the sphere by the double torus, and a general form for higher genera*, Journal of Combinatorial Theory, Series A **88** (1999), 259-275.
21. I. P. Goulden, D. M. Jackson, *The KP hierarchy, branched covers, and triangulations*, Advances in Mathematics **219** (2008), 932-951.
22. I. P. Goulden, D. M. Jackson, R. Vakil, *The Gromov-Witten potential of a point, Hurwitz numbers, and Hodge integrals*, Proceedings of the London Mathematical Society **83** (2001), 563-581.
23. I. P. Goulden, D. M. Jackson, R. Vakil, *Towards the geometry of double Hurwitz numbers*, Advances in Mathematics **198** (2005), 43-92.
24. D. J. Gross, W. Taylor IV, *Two-dimensional QCD is a string theory*, Nuclear Physics B **400** (1993), 181-208.
25. A. Guionnet, *First order asymptotics of matrix integrals; a rigorous approach towards the understanding of matrix models*, Communications in Mathematical Physics **244** (2004), 527-569.
26. A. Guionnet, *Large deviations and stochastic calculus for large random matrices*, Probability Surveys **1** (2004), 72-172.
27. A. Guionnet, *Random matrices and enumeration of maps*, Proceedings of the International Congress of Mathematicians, Madrid, Spain, 2006.
28. A. Guionnet, O. Zeitouni, *Large deviations asymptotics for spherical integrals*, Journal of Functional Analysis **188** (2002), 461-515.
29. P. Hanlon, R. P. Stanley, *A q-deformation of a trivial symmetric group action*, Transactions of the American Mathematical Society **350**(11) (1998), 4445-4459.
30. Harish-Chandra, *Differential operators on a semisimple Lie algebra*, American Journal of Mathematics **79** (1957), 87-120.
31. R. Howe, *Remarks on classical invariant theory*, Transactions of the American Mathematical Society **313**(2) (1989), 539-570.
32. A. Hurwitz, *Ueber Riemann'sche Flächen mit gegebenen Verzweigegungspunkten*, Mathematische Annalen **39** (1891), 1-60.
33. A. Hurwitz, *Ueber die Anzahl der Riemann'sch Flächen mit gegebenen Verzweigegungspunkten*, Mathematische Annalen **55** (1902), 53-66.
34. C. Itzykson, J.-B. Zuber, *The planar approximation. II*, Journal of Mathematical Physics **21**(3) (1980), 411-421.
35. P. Johnson, *Double Hurwitz numbers via the infinite wedge*, arXiv:1008.3266
36. A.-A. A. Jucys, *Symmetric polynomials and the center of the symmetric group ring*, Reports on Mathematical Physics **5**(1) (1974), 107-112.
37. M. E. Kazarian, S. K. Lando, *An algebro-geometric proof of Witten's conjecture*, Journal of the AMS **20**(4) (2007), 1079-1089.

38. S. Kerov, G. Olshanski, *Polynomial functions on the set of Young diagrams*, Comptes Rendus Acad. Sci. Paris Sér. I **319** (1994), 121-126.
39. M. Kontsevich, *Intersection theory on the moduli space of curves and the matrix Airy function*, Communications in Mathematical Physics **147** (1993), 1-23.
40. T. Y. Lam, *Representations of finite groups: a hundred years, part I*, Notices of the AMS **45**(3) (1998), 361-372.
41. S. Lando, A. Zvonkin, *Graphs on Surfaces and their Applications*, Encyclopedia of Mathematical Sciences, Volume **141**, 2004.
42. S. Matsumoto, J. Novak, *Jucys-Murphy elements and unitary matrix integrals*, arXiv:0905.1992
43. A. Matytsin, *On the large-N limit of the Itzykson-Zuber integral*, Nuclear Physics B **411** (1994), 805-820.
44. M. Mirzakhani, *Weil-Petersson volumes and intersection theory on the moduli space of curves*, Journal of the American Mathematical Society **20**(1) (2007), 1-23.
45. J. Novak, *Jucys-Murphy elements and the unitary Weingarten function*, Banach Center Publications **89** (2010), 231-235.
46. A. Okounkov, *Toda equations for Hurwitz numbers*, Mathematical Research Letters **7** (2000), 447-453.
47. A. Okounkov, *Infinite wedge and random partitions*, Selecta Mathematica, New Series **7** (2001), 57-81.
48. A. Okounkov, G. Olshanski, *Shifted Schur functions*, St. Petersburg Mathematics Journal **9** (1998), 239-300.
49. A. Okounkov, R. Pandharipande, *Gromov-Witten theory, Hurwitz numbers, and matrix models, I*, arXiv:0101147
50. A. Okounkov, R. Pandharipande, *Gromov-Witten theory, Hurwitz theory, and completed cycles*, Annals of Mathematics **163**(2) (2006), 517-560.
51. A. Okounkov, A. Vershik, *A new approach to the representation theory of the symmetric groups*, Selecta Mathematica **2**(4) (1996), 581-605.
52. G. Olshanski, *Plancherel averages: remarks on a paper by Stanley*, The Electronic Journal of Combinatorics **17** (2010), #R43.
53. R. Pandharipande, *The Toda equations and the Gromov-Witten theory of the Riemann sphere*, Letters in Mathematical Physics **53** (2000), 59-74.
54. V. Periwal, D. Shevitz, *Unitary-matrix models as exactly solvable string theories*, Physical Review Letters **64**(12) (1990), 1326-1329.
55. S. Samuel, $\mathbf{U}(N)$*-integrals,* $1/N$*, and the De Wit - 't Hooft anomalies*, Journal of Mathematical Physics **21**(12) (1980), 2695-2703.
56. S. Shadrin, M. Shapiro, A. Vainshtein, *On double Hurwitz numbers in genus zero*, Formal Power Series and Algebraic Combinatorics, Nankai University, Tianjin, 2007.
57. S. Shadrin, L. Spitz, D. Zvonkine, *On double Hurwitz numbers with completed cycles*, arXiv:1103.3120
58. R. P. Stanley, *Increasing and decreasing subsequences and their variants*, Proceedings of the International Congress of Mathematicians, Madrid, Spain, 2006.
59. G. Strang, Online lectures on linear algebra, MIT open courseware, Lecture 15, available at http://www.youtube.com
60. K. Takasaki, K. Ueno, *Toda lattice hierarchy*, Advanced Studies in Pure Mathematics **4** (1984), Group Representations and Systems of Differential Equations, 1-95.
61. G. 't Hooft, *A planar diagram theory for strong interactions*, Nuclear Physics B **72** (1974), 461-473.
62. W. T. Tutte, *A census of Hamiltonian polygons*, Canad J. Math. **14** (1962), 402 –417.
63. A. Varchenko, *Bilinear form of real configuration of hyperplanes*, Advances in Mathematics **97** (1993), 110-144.
64. E. Witten, *Quantum gravity and intersection theory on the moduli space of curves*, Surveys in Differential Geometry **1** (1991), 243-310.
65. D. Zagier, *Realizability of a model in infinite statistics*, Communications in Mathematical Physics **147** (1992), 199-210.
66. P. Zinn-Justin, *HCIZ integral and 2D Toda lattice hierarchy*, Nuclear Physics B **634** [FS] (2002), 417-432.

67. P. Zinn-Justin, J.-B. Zuber, *On some integrals over the $U(N)$ unitary group and their large N limit*, Journal of Physics A: Mathematical and General **36** (2003), 3173-3193.

Department of Combinatorics & Optimization. University of Waterloo, Canada
E-mail address: ipgoulden@uwaterloo.ca

Department of Combinatorics & Optimization. University of Waterloo, Canada
E-mail address: mguaypaq@uwaterloo.ca

Department of Combinatorics & Optimization. University of Waterloo, Canada
E-mail address: j2novak@uwaterloo.ca