

## CLASSICAL HURWITZ NUMBERS AND RELATED COMBINATORICS

BORIS DUBROVIN, DI YANG, AND DON ZAGIER

*To the memory of the extraordinary mathematician and man  
Vladimir Igorevich Arnold,  
with admiration*

**ABSTRACT.** We give a polynomial-time algorithm of computing the classical Hurwitz numbers  $H_{g,d}$ , which were defined by Hurwitz 125 years ago. We show that the generating series of  $H_{g,d}$  for any fixed  $g \geq 2$  lives in a certain subring of the ring of formal power series that we call the Lambert ring. We then define some analogous numbers appearing in enumerations of graphs, ribbon graphs, and in the intersection theory on moduli spaces of algebraic curves, such that their generating series belong to the same Lambert ring. Several asymptotics of these numbers (for large  $g$  or for large  $d$ ) are obtained.

2010 MATH. SUBJ. CLASS. Primary: 14N10; Secondary: 16T30, 53D45, 05A15.

KEY WORDS AND PHRASES. Hurwitz numbers, Lambert ring, Pandharipande's equation, enumerative geometry.

In 1891 Hurwitz [30] studied the number  $H_{g,d}$  of genus  $g \geq 0$  and degree  $d \geq 1$  coverings of the Riemann sphere with  $2g + 2d - 2$  fixed branch points and in particular found a closed formula for  $H_{g,d}$  for any fixed  $d$ . These Hurwitz numbers are now very famous and have been studied from many different points of view (matrix models, Gromov–Witten invariants, topological recursion, classical/quantum integrable systems, ...); see for example [4], [7], [10], [12], [27], [28], [42], [43]. In this paper we study their combinatorial properties, and compare them with some other enumerative problems.

Introduce the following generating series of  $H_{g,d}$

$$\mathcal{H} = \mathcal{H}(x, y) = \sum_{g \geq 0} \sum_{d \geq 1} h_{g,d} x^{2g+2d-2} y^d, \quad h_{g,d} := \frac{H_{g,d}}{(2g + 2d - 2)!}. \quad (1)$$

In the first part of the paper (Section 1), which is completely elementary, we will deduce from a recursion relation obtained by R. Pandharipande in [43] a much simpler one, which can be written compactly in terms of the generating series (1)

---

Received March 20, 2017; in revised form June 30, 2017.

in the following form

$$\begin{aligned} \mathcal{H}^{(3)} - \mathcal{H}^{(2)} &= 2\mathcal{H}^{(2)} \sum_{\ell=1}^{\infty} \frac{x^{2\ell}}{(2\ell)!} \mathcal{H}^{(2\ell+1)}, \\ \mathcal{H}^{(n)} &:= \left(y \frac{\partial}{\partial y}\right)^n \mathcal{H} = \sum_{g \geq 0} \sum_{d \geq 1} d^n h_{g,d} x^{2g+2d-2} y^d, \end{aligned} \tag{2}$$

and use it to give simpler derivations of some known and new properties of  $H_{g,d}$ . In particular:

P1. For any fixed  $d \geq 1$ , the function  $H_{g,d}$  is a linear combination of finitely many exponentials  $m^{2g+2d-2}$ ,  $1 \leq m \leq \binom{d}{2}$ , with coefficients in  $2d!^{-2}\mathbb{Z}$ . For example, for  $d = 6$

$$\begin{aligned} \frac{6!^2}{2} H_{g,6} &= 15^w - 36 \cdot 10^w + 25 \cdot 9^w - 225 \cdot 7^w + 700 \cdot 6^w - 720 \cdot 5^w \\ &\quad + 7200 \cdot 4^w - 15200 \cdot 3^w - 34200 \cdot 2^w + 163575, \end{aligned}$$

where  $w = 2g + 10$ . (This example was already given by Hurwitz). In particular, we have the asymptotic formula

$$H_{g,d} \sim \frac{2}{d!^2} \binom{d}{2}^{2g+2d-2} \quad \text{as } g \rightarrow \infty. \tag{3}$$

The exponential polynomials  $H_{*,d}$  can be computed recursively in time polynomial in  $d$ .

P2. For any fixed  $g \geq 0$ , the function  $d! h_{g,d}$  belongs to the rank 2 module over  $\mathbb{Q}[d]$  generated by the two integer-valued functions

$$d \mapsto 2d^{d-3} \quad \text{and} \quad d \mapsto (d-1)! \sum_{r=0}^{d-1} \frac{d^r}{r!}$$

and  $h_{g,d}$  itself is a linear combination of  $h_{0,d}$  and  $h_{1,d}$  with polynomial coefficients, e.g.,

$$\begin{aligned} h_{2,d} &= \frac{7d^4(d-1)}{4320} h_{0,d} - \frac{5d^2}{72} h_{1,d}, \\ h_{3,d} &= -\frac{d^4(d-1)(99845d^2 - 454d + 24)}{1045094400} h_{0,d} \\ &\quad + \frac{d^2(128625d^3 + 546700d^2 - 980d + 48)}{174182400} h_{1,d}. \end{aligned} \tag{4}$$

Moreover, the series  $\mathcal{H}_g^{(n)}(z) = \sum_{d \geq 1} h_{g,d} d^n z^d$  is a polynomial with rational coefficients in  $\mathcal{H}_0^{(3)}(z)$  for all  $g$  and  $n$  satisfying the stability condition  $2g - 2 + n > 0$ , and there is a polynomial time (in  $g$  and  $n$ ) recursive algorithm to compute these polynomials.

P3. The asymptotic behavior of  $H_{g,d}$  for  $g \geq 0$  fixed is given by

$$H_{g,d} \sim \frac{c_g \sqrt{\pi/2}}{(24\sqrt{2})^g \Gamma(\frac{5g-1}{2})} \left(\frac{4}{e}\right)^d d^{2d-5+9g/2} \quad \text{as } d \rightarrow \infty \tag{5}$$

or equivalently

$$h_{g,d} \sim \frac{c_g \sqrt{2}}{(96\sqrt{2})^g \Gamma(\frac{5g-1}{2})} d^{(5g-7)/2} e^d \quad \text{as } d \rightarrow \infty, \tag{6}$$

where the numbers

$$c_0 = -1, \quad c_1 = 2, \quad c_2 = 98, \quad c_3 = 19600, \quad c_4 = 8824802, \quad c_5 = 7061762400, \quad \dots$$

are given by the recursion

$$c_g = 50(g-1)^2 c_{g-1} + \frac{1}{2} \sum_{h=2}^{g-2} c_h c_{g-h} \quad (g \geq 3) \tag{7}$$

or equivalently by the statement that the generating series

$$U = U(X) = \sum_{g \geq 0} c_g X^{\frac{1}{2} - \frac{5g}{2}}$$

satisfies the Painlevé I equation

$$\frac{d^2 U}{dX^2} + \frac{1}{16} U^2 - \frac{1}{16} X = 0. \tag{8}$$

We indicate briefly which of these results were known and what is new. The first statement of P1 is due to Hurwitz, as already mentioned, and the asymptotic formula (3) follows immediately, but the last statement is new. (Hurwitz’s formulas for  $H_{*,d}$  involve summations over all permutations or all partitions of  $d$  and therefore contain an exponentially growing number of terms.) The first statement in P2 seems not to be in the literature. The second statement, that every  $h_{g,*}$  is a linear combination of  $h_{0,*}$  and  $h_{1,*}$  with polynomial coefficients is implicit in the work of Goulden, Jackson and Vakil [28]. However, the explicit computations were done there only up to genus 3 (Examples 4.1–4.3), and actual formulas for higher genera would be almost impossible to obtain by their approach, which is based on difficult computations of Hodge integrals, whereas with our polynomial-time algorithm given here one can easily compute up to much larger values of the genus. The formula for  $h_{3,d}$  given in [28] expresses it as a linear combination with polynomial coefficients of  $h_{1,d}$  and  $h_{2,d}$  (one sees from (4) that this is possible, since the coefficient of  $h_{0,d}$  in  $h_{3,d}$  is divisible by the coefficient of  $h_{0,d}$  in  $h_{2,d}$ ) and a remark given by the authors in this context (“This is to be expected to persist for  $g \geq 2$ ,” p. 578) seems to suggest that more generally  $h_{g,d}$  is a combination with coefficients in  $\mathbb{Q}[d]$  of  $h_{g-1,d}$  and  $h_{g-2,d}$  for any  $g$ , but in fact this fails already for  $g = 4$ . (Of course  $h_{g,d}$  is always a combination of  $h_{g-1,d}$  and  $h_{g-2,d}$  with coefficients in  $\mathbb{Q}(d)$ , but in general the coefficients will have denominators.) Finally, a result equivalent to P3 was proved by Caporaso et al. [6] as a consequence of the Ekedahl–Lando–Shapiro–Vainshtein formula [21] along with Itzykson–Zuber’s result [32] on the asymptotics of certain Gromov–Witten invariants of a point, whereas we give a direct proof using just the recursion.

We also remark that to obtain the results of P2 we study two simple spaces of sequences of rational numbers (the “Lambert space” and “extended Lambert space”) that contain the sequences  $d \mapsto d! h_{g,d}$  with  $g$  fixed and many other interesting counting functions. Related spaces have appeared in the literature many

times, most notably in [46], but our way of studying these spaces and several of the specific results that we prove about them are new.

The second part of this paper (Section 2), describes a different method of computing  $H_{g,d}$  for fixed  $g$  based on the Dubrovin–Zhang (DZ) approach [18]. It was shown by Pandharipande [43] that  $H_{g,d}$  are special Gromov–Witten (GW) invariants of the complex projective line  $\mathbb{P}^1$ . In [9], [18], [11], the first named author and Y. Zhang developed a powerful method of computing GW invariants of any smooth projective variety with semisimple quantum cohomology. The quantum cohomology of  $\mathbb{P}^1$  is semisimple. Indeed, the small quantum cohomology of  $\mathbb{P}^1$  at the origin is already semisimple:

$$1 \star 1 = 1, \quad 1 \star \omega = \omega \star 1 = \omega, \quad \omega \star \omega = 1, \quad (9)$$

where  $1 \in H^0(\mathbb{P}^1; \mathbb{Q})$ ,  $\omega \in H^2(\mathbb{P}^1; \mathbb{Q})$  is the Poincaré dual of a point, and  $\star$  denotes the quantum product. Hence one can apply the DZ approach to the computation of  $H_{g,d}$ . We mention that, according to [9], [18], the multiplication table (9) together with the Poincaré pairing on  $H^*(\mathbb{P}^1; \mathbb{Q})$  contain complete information on all GW invariants of  $\mathbb{P}^1$ .

In the third part of this paper (Section 3), we investigate *three* more models in enumerative geometry that share similar properties with Hurwitz’s counting problem. The first one is the enumeration of ordinary graphs. The second one is the enumeration of ribbon graphs (aka Grothendieck’s dessins d’enfants) [3], [29], [36]. Although closed formulas for the enumeration of ribbon graphs with *one* and *two* vertices had been obtained in [29] and [38], respectively, it is only recently that two efficient algorithms for the ribbon graph enumeration (one is efficient for large genus, the other is efficient for large number of vertices) have been developed by two of the authors [16], [17]. The algorithm in [17] observes an interesting relationship between Hodge integrals and matrix integrals called the *Hodge–GUE correspondence*. The validity of this algorithm has been proved very recently in [14]. The third one is the study of intersection numbers (of the so-called  $\psi$ -classes) over the Deligne–Mumford moduli space  $\overline{\mathcal{M}}_{g,n}$  of stable algebraic curves of genus  $g$  with  $n$  distinct marked points along with nonlinear Hodge classes of the most general type [39], [23], [13]. The comparisons between these three models and Hurwitz’s problem with  $g \geq 2$  are summarized in the following list:

	order of the most singular term	characterization of the top coefficient
Ordinary graphs	$3g - 3$	Riccati equation
Ribbon graphs	$4g - 4$	not known
Hodge integrals	$5g - 5$	Painlevé I equation
Hurwitz numbers	$5g - 5$	Painlevé I equation

Here, for Hurwitz’s problem,  $g$  is the genus of the upper Riemann surface of a covering; for Hodge integrals,  $g$  is the genus of the stable curves that form the moduli space; for an ordinary graph,  $g$  is the number of loops of the graph; for a ribbon graph,  $g$  is defined as the smallest genus of a Riemann surface into which the ribbon graph can be embedded, while the words “most singular term” and “the

top coefficient” refer to the asymptotic growth of the elements of the sequence in question for fixed  $g$ .

1. ELEMENTARY THEORY OF HURWITZ NUMBERS

In this section we derive the main combinatorial properties of the Hurwitz numbers from a completely elementary point of view, not involving any geometric considerations. Our main tool is a quadratic recursion that we give in Section 1.1. The discussions of the Hurwitz numbers for fixed  $d$  or fixed  $g$  are given in Sections 1.2 and 1.3, respectively, while Sections 1.4 and 1.5 discuss the Lambert space and the asymptotic properties of the Hurwitz numbers.

**1.1. A quadratic recursion for the Hurwitz numbers.** It was conjectured by Pandharipande [43] and later proved by Okounkov [40] that the generating series  $\mathcal{H}$  defined in (1) satisfies the following differential-functional equation

$$D^2\mathcal{H}(x, y) = y e^{\mathcal{H}(x, ye^x) - 2\mathcal{H}(x, y) + \mathcal{H}(x, ye^{-x})}, \quad D := y \frac{\partial}{\partial y}. \tag{10}$$

Expanding  $\mathcal{H}(x, ye^{\pm x})$  by Taylor’s theorem and comparing the coefficients of both sides, one can rewrite (10) equivalently as a recursion for the numbers  $h_{g,d} = H_{g,d}/(2g + 2d - 2)!$  (page 64 of [43]):

$$h_{g,d} = \frac{1}{d^2} \sum_{\ell=0}^{\infty} \frac{2^\ell}{\ell!} \sum_{\substack{d_1, \dots, d_\ell \geq 1 \\ \sum d_i = d-1}} \sum_{\substack{g_1, \dots, g_\ell \geq 0 \\ k_1, \dots, k_\ell \geq 1 \\ \sum (g_i + k_i - 1) = g}} \prod_{i=1}^{\ell} \frac{d^{2k_i}}{(2k_i)!} h_{g_i, d_i}. \tag{11}$$

Our first observation is that both (10) and (11) can be replaced by much simpler quadratic equations.

**Proposition 1.** *The generating series  $\mathcal{H}$  satisfies the differential-functional equation*

$$D^3\mathcal{H}(x, y) - D^2\mathcal{H}(x, y) = D^2\mathcal{H}(x, y) (D\mathcal{H}(x, ye^x) - 2D\mathcal{H}(x, y) + D\mathcal{H}(x, ye^{-x})). \tag{12}$$

Equivalently, the numbers  $h_{g,d} = H_{g,d}/(2g + 2d - 2)!$  are given recursively by

$$h_{g,d} = \frac{2}{d^2(d-1)} \sum_{\substack{d_1, d_2 \geq 1 \\ d_1 + d_2 = d}} \sum_{\substack{g_1, g_2 \geq 0, \ell \geq 1 \\ g_1 + g_2 + \ell = g+1}} \frac{d_1^{2\ell+1} d_2^2}{(2\ell)!} h_{g_1, d_1} h_{g_2, d_2}. \tag{13}$$

*Proof.* Equation (12) follows from (10) by applying  $D$  to both sides, and (13) from (12). □

The quadratic recursion (13) is useful not only for theoretical considerations, as we will see in the rest of this section, but also for numerical purposes, since the number of terms on the right grows only polynomially (like  $O(g^2d)$ ) rather than exponentially in the arguments  $g$  and  $d$ .

**1.2. Hurwitz numbers for fixed degree.** Our first application of Proposition 1 is to study the numbers  $h_{g,d}$  for fixed  $d$  by giving a recursive formula for the generating functions

$$C_d(x) := \sum_{g \geq 0} h_{g,d} x^{2g-2+2d} \quad (d \geq 1).$$

This also gives a convenient algorithm for computing the Hurwitz numbers, since the recursion for  $C_d$  involves only  $O(d)$  terms, so that we can easily calculate up to fairly large values of  $d$ .

To state the recursion we will use a sequence of polynomials  $P_k = P_k(s)$ , with initial values

$$P_1 = s, \quad P_2 = s^2 + 4s, \quad P_3 = s^3 + 6s^2 + 9s, \quad P_4 = s^4 + 8s^3 + 20s^2 + 16s, \quad \dots,$$

that can be defined either by  $P_k(s) = 2T_k\left(\frac{s+2}{2}\right) - 2$  ( $T_k$  = Chebyshev polynomial of the first kind) or else recursively by  $P_{k+1} = (s + 2)P_k - P_{k-1} + 2s$ , or else in closed form as  $\sum_{i=1}^k \left(\binom{k+i}{2i} + \binom{k+i-1}{2i}\right) s^i$ . (The equivalence of these three definitions is elementary.) Then we have:

**Theorem 1.** *The generating series  $C_d(x) \in \mathbb{Q}[[x]]$  is given for each  $d \geq 1$  by*

$$C_d(x) = \frac{1}{d} \gamma_d(4 \sinh^2(x/2)), \tag{14}$$

where  $\gamma_d(s)$  is a polynomial of degree  $\binom{d}{2}$  in  $s$  defined inductively by

$$\gamma_d(s) = \frac{1}{d^2 - d} \sum_{k=1}^{d-1} (d - k) P_k(s) \gamma_k(s) \gamma_{d-k}(s) \quad (d \geq 2) \tag{15}$$

together with the initial condition  $\gamma_1(s) = 1$ , the first values being

$$\gamma_1(s) = 1, \quad \gamma_2(s) = \frac{s}{2!1!}, \quad \gamma_3(s) = \frac{s^3 + 6s^2}{3!2!}, \quad \gamma_4(s) = \frac{s^6 + 12s^5 + 54s^4 + 96s^3}{4!3!}, \quad \dots$$

*Proof.* Since  $\mathcal{H} = \sum_d C_d(x) y^d$ , we have

$$D\mathcal{H}(x, y e^x) - 2D\mathcal{H}(x, y) + D\mathcal{H}(x, y e^{-x}) = \sum_{d \geq 1} d C_d(x) (e^{dx} - 2 + e^{-dx}) y^d. \tag{16}$$

Using (16) and comparing coefficients of  $y^d$  of the both sides of (12) we find

$$(d^3 - d^2) C_d(x) = \sum_{k=1}^{d-1} k(d - k)^2 (e^{kx} - 2 + e^{-kx}) C_k(x) C_{d-k}(x). \tag{17}$$

After changing variables from  $x$  to  $s = 4 \sinh^2(x/2)$  (and also multiplying  $C_d$  by  $d$  to simplify the recursion slightly), this takes the form (15) with  $P_k(s)$  defined by  $P_k(x - 2 + x^{-1}) = x^k - 2 + x^{-k}$ , which is equivalent to each of the definitions of  $P_k$  given just before the theorem.  $\square$

**Remark.** Although equation (12) in Proposition 1 follows easily from Pandharipande’s equation, it changes considerably the form of the recursion for  $H_{g,d}$ , and significantly reduces the time complexity. This is actually an unexpected phenomenon, essentially due to the nonlinearity of Pandharipande’s original equation (10).

To the best of our knowledge, the algorithm described in Theorem 1 is the first one in the literature with the *polynomial* time complexity for computing  $H_{g,d}$ . See in Sections 1.5 and 1.3 for more discussions as well as applications.

The following corollary is a more precise version of the statement P1 of the introduction.

**Corollary** (Hurwitz 1891). *The numbers  $H_{g,d}$  for fixed  $d$  have the form*

$$H_{g,d} = \frac{2}{d!^2} \sum_{1 \leq m \leq \binom{d}{2}} b_{d,m} m^{2g+2d-2}, \tag{18}$$

where  $b_{d,m}$  are integers with  $b_{d,\binom{d}{2}} = 1$  and  $b_{d,m} = 0$  for  $\binom{d-1}{2} < m < \binom{d}{2}$ .

*Proof.* From Proposition 1 or from the recursion (17) we see that  $C_d(x)$  is a Laurent polynomial in  $e^x$  and is even in  $x$ . This implies a formula of the form (18) with  $b_{d,m} \in \mathbb{Q}$ . The further properties (integrality and values for  $m > \binom{d-1}{2}$ ) are also easily deduced from (17).  $\square$

We end this subsection by discussing another approach to calculating the numbers  $H_{g,d}$  and proving the corollary above that was discovered by Hurwitz himself in a second paper [31] in 1901. Let  $H_{k,d}^*$  be the weighted (by inverse number of automorphisms) number of (not necessarily connected) coverings of  $\mathbb{P}^1$  of degree  $d$  with  $k$  simple ramification points, and introduce the corresponding generating series

$$Z_{\mathcal{H}} = Z_{\mathcal{H}}(x, y) := \sum_{k, d \geq 0} \frac{H_{k,d}^*}{k!} x^k y^d =: \sum_{d \geq 0} C_d^*(x) y^d. \tag{19}$$

Note that the connected coverings in this setting have genus given by  $k = 2g + 2d - 2$ , so by a standard argument we know that  $Z_{\mathcal{H}}$  is related to the generating series (1) by

$$Z_{\mathcal{H}}(x, y) = e^{\mathcal{H}(x,y)}. \tag{20}$$

Hurwitz showed that  $d! H_{k,d}^*$  is the number of homomorphisms from the group  $\pi_1(S^2 \setminus \{P_1, \dots, P_k\})$  to  $S_d$  for which the image of the generator at each puncture is a transposition, i.e., to the number of ordered  $k$ -tuples of transpositions in  $S_d$  with product 1. The famous Frobenius formula then gives

$$H_{k,d}^* = \frac{1}{d!^2} \sum_{\pi} (\dim \pi)^2 \nu(\pi)^k,$$

where the sum is over all irreducible representations  $\pi$  of  $S_d$  and  $\nu(\pi)$  is the value (scalar, by Schur’s lemma) of the central element  $\sum_{i < j} [(ij)]$  in  $\mathbb{Z}[S_d]$  on  $\pi$ . (See also [41], [12].) Hence  $C_d^*(x) = \frac{1}{d!^2} \sum_{\pi} (\dim \pi)^2 e^{\nu(\pi)x}$ , which already has the desired property of being a Laurent polynomial in  $e^x$  (and also even in  $x$ , as of course it must be from its definition, because  $\nu(\pi^\vee) = -\nu(\pi)$  for all  $\pi$ ) and hence gives a direct proof of the formula (14) for some polynomial  $\gamma_d(s)$ . Hurwitz [31] found explicit expressions for  $\dim \pi$  and  $\nu(\pi)$  in terms of partitions and gave the following

beautiful closed formula for  $C_d^*(x)$

$$C_d^*(x) = \sum_{\substack{0 \leq n_1 < n_2 < \dots < n_d \\ n_1 + \dots + n_d = d(d+1)/2}} \left( \frac{\prod_{i < j} (n_i - n_j)}{\prod_i n_i!} \right)^2 e^{(\frac{1}{2} \sum_{i=1}^d n_i^2 - \frac{1}{12} d(d+5)(2d-1))x}, \quad (21)$$

as well as the initial values

$$\begin{aligned} \frac{2!^2}{2} C_2^*(x) &= \cosh x, \\ \frac{3!^2}{2} C_3^*(x) &= \cosh(3x) + 2, \\ \frac{4!^2}{2} C_4^*(x) &= \cosh(6x) + 9 \cosh(2x) + 2, \\ \frac{5!^2}{2} C_5^*(x) &= \cosh(10x) + 16 \cosh(5x) + 25 \cosh(2x) + 18. \end{aligned}$$

We now observe that  $Z_{\mathcal{H}}$ , like  $\mathcal{H}$  itself, satisfies a quadratic differential-functional equation. Indeed, equation (20) and Pandharipande’s equation (10) imply

$$Z_{\mathcal{H}}(x, y) D^2 Z_{\mathcal{H}}(x, y) - D Z_{\mathcal{H}}(x, y)^2 = y Z_{\mathcal{H}}(x, y e^x) Z_{\mathcal{H}}(x, y e^{-x}). \quad (22)$$

Substituting (19) and comparing coefficients of powers of  $y$  we arrive at

**Theorem 2.** *The functions  $C_d^*(x)$ ,  $d \geq 1$ , are given recursively by*

$$\begin{aligned} d^2 C_d^*(x) &= \sum_{\substack{d_1, d_2 \geq 1 \\ d_1 + d_2 = d}} (d_1 - d_2)^2 C_{d_1}^*(x) C_{d_2}^*(x) \\ &\quad + \sum_{\substack{d_1, d_2 \geq 0 \\ d_1 + d_2 = d-1}} e^{(d_1 - d_2)x} C_{d_1}^*(x) C_{d_2}^*(x), \quad d \geq 1, \end{aligned} \quad (23)$$

together with the initial data  $C_0^*(x) = 1$ .

Note that (21) has a number of terms growing more than polynomially in  $d$  (specifically, like the number of partitions of  $d$ ), whereas (23) lets one compute  $C_d^* \in \mathbb{Q}[\cosh x]$  for all  $d \leq D$  in  $O(D^2)$  steps. It would be very nice if one could use Hurwitz’s explicit formula (21) to give a purely elementary proof of the quadratic recursion (23), and hence also of Pandharipande’s original recursion (10).

**1.3. Hurwitz numbers for fixed genus.** In the previous subsection we studied the structure of the generating function of the Hurwitz numbers  $H_{g,d}$  for fixed  $d$ . We now turn to the complementary case when  $g$  is fixed, i.e., we want to describe the generating function

$$\mathcal{H}_g = \mathcal{H}_g(z) := \sum_{d=1}^{\infty} h_{g,d} z^d \quad (24)$$

for every  $g \geq 0$ , and in particular to prove the statements given in P2 of the introduction.

We set  $z = x^2y$  and consider  $x$  as fixed, so that the operator  $D$  defined in (10) can also be written as  $z \frac{d}{dz}$ . From the definitions (1) and (24) we have

$$\mathcal{H} = \sum_{g \geq 0} \mathcal{H}_g(x^2y) x^{2g-2} = \sum_{g \geq 0} \mathcal{H}_g(z) x^{2g-2},$$

$$\mathcal{H}^{(n)} := D^n \mathcal{H} = \sum_{g \geq 0} \mathcal{H}_g^{(n)}(z) x^{2g-2} \quad (n \geq 0).$$

Applying to this our quadratic differential equation (12) and comparing coefficients of  $x^{2g-2}$  on both sides, we obtain the differential-recursive equation

$$\mathcal{H}_g^{(3)} - \mathcal{H}_g^{(2)} = \sum_{\substack{g_1, g_2 \geq 0, \ell \geq 1 \\ g_1 + g_2 + \ell = g + 1}} \frac{2}{(2\ell)!} \mathcal{H}_{g_1}^{(2)} \mathcal{H}_{g_2}^{(2\ell+1)} \tag{25}$$

for the functions  $\mathcal{H}_g$ , from which all of the desired properties will follow. Notice that this equation is autonomous, i.e., it does not contain the independent variable  $z$ . This will be crucial for us since it means that we can freely replace  $z$  by any equivalent variable (i.e., one given by an invertible power series in  $z$ ) and write  $D$  with respect to the new variable without changing (25).

We first consider (25) for small values of  $g$ . For  $g = 0$  it becomes  $\mathcal{H}_0^{(3)} - \mathcal{H}_0^{(2)} = \mathcal{H}_0^{(2)} \mathcal{H}_0^{(3)}$ , which integrated once gives  $\mathcal{H}_0^{(2)} - \mathcal{H}_0^{(1)} = \frac{1}{2} (\mathcal{H}_0^{(2)})^2$ . (Notice that there is no constant of integration since all of our power series have constant term 0 and  $D$  is invertible on such power series.) So if we set

$$T = \mathcal{H}_0^{(2)}(z) = z + z^2 + \frac{3}{2} z^3 + \frac{8}{3} z^4 + \dots, \tag{26}$$

then  $\mathcal{H}_0^{(1)} = T - \frac{1}{2} T^2$  and hence  $D(T - \frac{1}{2} T^2) = T$  or  $\frac{1}{z} \frac{dz}{dT} = \frac{1-T}{T}$ , which integrates immediately to

$$z = T e^{-T}. \tag{27}$$

Equation (27) expresses  $z$  as a power series in  $T$  and hence conversely determines  $T$  as a power series in  $z$ . This is the so-called *Lambert function*, whose well-known expansion is given by

$$T = \sum_{d=1}^{\infty} \frac{d^{d-1}}{d!} z^d. \tag{28}$$

Together with the identification  $T = \mathcal{H}_0^{(2)}(z)$  this gives the genus 0 Hurwitz numbers as  $h_{0,d} = d^{d-3}/d!$ . To write the generating function  $\mathcal{H}_0$  itself, rather than its second derivative, as a function of  $T$ , we note that the relation  $D(T) = T/(1 - T)$  implies the basic formula

$$D := z \frac{d}{dz} = \frac{T}{1 - T} \frac{d}{dT} \tag{29}$$

for the differential operator  $D$  in terms of  $T$ , and using this we can integrate or differentiate any expression in  $T$  as many times as we want, obtaining in particular

$$\begin{aligned} \mathcal{H}_0 &= T - \frac{3}{4}T^2 + \frac{1}{6}T^3, & \mathcal{H}_0^{(1)} &= T - \frac{1}{2}T^2, & \mathcal{H}_0^{(2)} &= T, \\ \mathcal{H}_0^{(3)} &= \frac{T}{1-T}, & \mathcal{H}_0^{(4)} &= \frac{T}{(1-T)^3}, & \mathcal{H}_0^{(5)} &= \frac{T+2T^2}{(1-T)^5}, \quad \dots \end{aligned} \tag{30}$$

We next consider  $g = 1$ . For this case (25) becomes

$$\mathcal{H}_1^{(3)} - \mathcal{H}_1^{(2)} = \frac{1}{12} \mathcal{H}_0^{(2)} \mathcal{H}_0^{(5)} + \mathcal{H}_0^{(2)} \mathcal{H}_1^{(3)} + \mathcal{H}_1^{(2)} \mathcal{H}_0^{(3)} = D \left( \frac{1}{24} \frac{T+T^2}{(1-T)^3} + T \mathcal{H}_1^{(2)} \right),$$

where for the second equality we have used (29) and the values in (30). Integrating once gives

$$\left( T \frac{d}{dT} - 1 \right) \mathcal{H}_1^{(1)} = [(1-T)D - 1] \mathcal{H}_1^{(1)} = \frac{1}{24} \frac{T+T^2}{(1-T)^3}, \tag{31}$$

which is easily integrated twice more to give

$$\mathcal{H}_1^{(1)} = \frac{1}{24} \frac{T^2}{(1-T)^2}, \quad \mathcal{H}_1 = -\frac{T + \log(1-T)}{24}. \tag{32}$$

Continuing in this way, we can find each power series  $\mathcal{H}_g(z)$  in terms of  $T$ , the next two cases being

$$\mathcal{H}_2 = \frac{T^2 + 6T^3}{1440(1-T)^5}, \quad \mathcal{H}_3 = \frac{9T^2 + 548T^3 + 3482T^4 + 3816T^5 + 720T^6}{725760(1-T)^{10}}. \tag{33}$$

But it is not obvious that the integration process always works and that each higher  $\mathcal{H}_g$  is a polynomial in  $1/(1-T)$ . This is part of the content of the following theorem.

**Theorem 3.** *Let  $z$  and  $T$  be variables related by (27). Then the generating series  $\mathcal{H}_g$  for all  $g \geq 0$  as power series in either  $z$  or  $T$  are uniquely determined by the recursive differential equation (25), where  $\mathcal{H}_g^{(n)} := D^n \mathcal{H}_g$  with  $D$  as in (29). They are always elementary functions of  $T$ , with  $\mathcal{H}_g^{(n)} \in \mathbb{Q}[1/(1-T)]$  for all  $g, n \geq 0$  satisfying the stability condition  $2g+n-2 > 0$ . Explicitly,  $\mathcal{H}_g$  is given by (30), (32) and (33) for  $g \leq 3$  and by an expression of the form*

$$\mathcal{H}_g = \sum_{i=2g-2}^{5g-5} \frac{\kappa_{g,i}}{(1-T)^i} \tag{34}$$

for all  $g \geq 2$ , with top and bottom coefficients given by

$$\kappa_{g,2g-2} = \frac{B_{2g}}{2g(2g-2)}, \quad \kappa_{g,5g-5} = \frac{24^{-g} c_g}{(5g-3)(5g-5)}, \tag{35}$$

where  $B_n$  denotes the  $n^{\text{th}}$  Bernoulli number and the  $c_g$  are the numbers defined by (7) or (8).

*Proof.* The recursive differential equation (25) is third-order in  $\mathcal{H}_g$ , so we have to integrate three times. Symmetrizing (25) and noting that  $A^{(2)}B^{(2m+1)} + A^{(2m+1)}B^{(2)}$

for any two functions  $A$  and  $B$  is the derivative of  $A^{(2)}B^{(2m)} - A^{(3)}B^{(2m-1)} + \dots + A^{(2m)}B^{(2)}$ , we can integrate (25) once to get

$$\mathcal{H}_g^{(2)} - \mathcal{H}_g^{(1)} = \sum_{\substack{g_1, g_2 \geq 0, n_1, n_2 \geq 2 \\ 2g_1 + 2g_2 + n_1 + n_2 = 2g + 4}} \frac{(-1)^{n_1}}{(n_1 + n_2 - 2)!} \mathcal{H}_{g_1}^{(n_1)} \mathcal{H}_{g_2}^{(n_2)}. \tag{36}$$

(As before, there is no constant of integration because all of our power series have constant term 0.) Separating out the terms involving  $\mathcal{H}_g$ , we can rewrite this as

$$\left(T \frac{d}{dT} - 1\right) \mathcal{H}_g^{(1)} = \sum_{\substack{0 \leq g_1, g_2 \leq g-1, n_1, n_2 \geq 2 \\ 2g_1 + 2g_2 + n_1 + n_2 = 2g + 4}} \frac{(-1)^{n_1}}{(n_1 + n_2 - 2)!} \mathcal{H}_{g_1}^{(n_1)} \mathcal{H}_{g_2}^{(n_2)}, \tag{37}$$

just as we did for the special case  $g = 1$ . To see that  $\mathcal{H}_g$  is a polynomial in  $(1 - T)^{-1}$ , we must show that the right-hand side of (37), which by induction on  $g$  is such a polynomial, is in the image of the operator  $(T \frac{d}{dT} - 1)D$  from  $\mathbb{Q}[(1 - T)^{-1}]$  to itself. This operator sends  $(1 - T)^{-n}$  to  $n(n + 2)T(1 - T)^{-n-3}$ , so its image is  $T(1 - T)^{-3}\mathbb{Q}[(1 - T)^{-1}]$ , i.e., the set of polynomials in  $(1 - T)^{-1}$  that vanish at  $T = 0$  and are  $O(T^{-2})$  as  $T \rightarrow \infty$ . The right-hand side of (37), which we abbreviate to  $R_g$ , obviously has the first property, since every series  $\mathcal{H}_g^{(n)}$  has zero constant term. For the second, we note that the first of equations (35), together with the special cases (30) and (32), implies that

$$\mathcal{H}_g^{(n)} \sim \frac{(2g - 1)B_{2g}}{(2g)!} \frac{(2g + n - 3)!}{T^{2g+n-2}} \text{ as } T \rightarrow \infty \tag{38}$$

for all  $(g, n)$  satisfying  $2g + n \geq 4$ . Applying this formula inductively to each factor  $\mathcal{H}_{g_i}^{(n_i)}$  in (37) (since  $g_i < g$ ), we find that every summand in  $R_g$  is  $O(T^{-2g})$  unless one of  $(g_1, n_1)$  or  $(g_2, n_2)$  is equal to  $(0, 2)$  or  $(0, 3)$ . (They cannot both be, since  $g > 1$ .) In these special cases (38) must be replaced by  $\mathcal{H}_0^{(2)} = T$ ,  $\mathcal{H}_0^{(3)} = -1 + O(T^{-1})$ , and a short calculation then gives

$$R_g = 4g(2g - 2)! \sum_{g_1=0}^{g-1} \frac{(2g_1 - 1)B_{2g_1}}{(2g_1)!(2g - 2g_1 + 2)!} T^{-2g+1} + O(T^{-2g}) = -\frac{B_{2g}}{T^{2g-1}} + O(T^{-2g})$$

as  $T \rightarrow \infty$ , where the second equality is a consequence of the standard recursion for Bernoulli numbers. This shows that  $R_g = O(T^{-2})$  as required, completing the proof that (37) can be integrated twice more to give  $\mathcal{H}_g \in \mathbb{Q}[(1 - T)^{-1}]$ , and also gives the desired asymptotics of  $\mathcal{H}_g(T)$  as  $T \rightarrow \infty$ , completing the proof of the first of equations (35) by induction. For the second one we proceed analogously, noting first that the second of equations (35), together with the special cases (30) and (32), gives

$$\mathcal{H}_g^{(n)} \sim \frac{c_g}{24g} \frac{2^{n-2} \left(\frac{5g-1}{2}\right)_{n-2}}{(1 - T)^{5g-5+2n}} \text{ as } T \rightarrow 1$$

for all  $g \geq 0$  and  $n \geq 2$  except  $(g, n) = (0, 2)$ . (Here  $(x)_m = x(x + 1) \dots (x + m - 1)$  is the  $m$ th Pochhammer symbol.) Since  $(5g_1 - 5 + 2n_1) + (5g_2 - 5 + 2n_2) \geq 5g - 2$  in (37), with equality only if  $n_1 = n_2 = 2$ , we find from this that  $R_g \sim C(1 - T)^{-5g+2}$  as  $T \rightarrow 1$ , where  $C$  is the sum of the contributions to (37) from  $(g_1, g_2, n_1, n_2)$

of the form  $(g_1, g - g_1, 2, 2)$ ,  $(g - 1, 0, 4, 2)$  or  $(0, g - 1, 2, 4)$ , and is then equal to  $24^{-g}c_g$  by a short calculation using the recursion (7), completing the proof by induction of also the second of equations (35).  $\square$

**Remark.** Equations equivalent to (25) were known [28] for  $g \leq 1$ , but to the best of our knowledge this short and elegant equation is new for general  $g$ . We expect that equation (25) or (36) have geometric meanings for any  $g$ . This would suggest a new and geometric proof of the Toda conjecture [42], [19], [7] for the GW invariants of  $\mathbb{P}^1$  in the stationary sector. Another essential statement of Theorem 3, namely the fact that  $\mathcal{H}_g^{(n)}$  is a polynomial in  $1/(1-T)$  whenever  $2g+n-2 > 0$ , and that  $\mathcal{H}_g$  for  $g \geq 2$  has the form (34), are not new results, but were already proved by Goulden, Jackson and Vakil [28] using the celebrated Ekedahl–Lando–Shapiro–Vainshtein (ELSV) formula [21] giving a correspondence between Hurwitz numbers and linear Hodge integrals. Our object in this section was to give an elementary proof using only the Pandharipande recursion for the Hurwitz numbers (or the easier quadratic recursion that it implies), and also to give explicit formulas for the top and bottom coefficients. In fact the formula for the top coefficient was also known, since it was already observed by Caporaso–Griguolo–Mariño–Pasquetti–Seminara [6] that this coefficient can be written in terms of a certain Hodge integral that had in fact already been computed by Itzykson–Zuber [32]. We explain in a few words how this works. The ELSV formula is

$$h_{g,d} = \frac{1}{d!} \int_{\overline{\mathcal{M}}_{g,d}} \frac{1 - \lambda_1 + \lambda_2 - \cdots + (-1)^g \lambda_g}{\prod_{p=1}^d (1 - \psi_p)} = \frac{1}{d!} \sum_{\substack{0 \leq k \leq g, k_1, \dots, k_d \geq 0 \\ k + \sum k_i = 3g - 3 + d}} (-1)^k \int_{\overline{\mathcal{M}}_{g,d}} \lambda_k \psi_1^{k_1} \cdots \psi_d^{k_d}, \quad (39)$$

where  $\overline{\mathcal{M}}_{g,d}$  denotes the Deligne–Mumford moduli space of stable algebraic curves of genus  $g$  with  $d$  distinct marked points,  $\psi_p$  the first Chern class of the  $p$ -th tautological line bundle on  $\overline{\mathcal{M}}_{g,d}$ , and  $\lambda_k$  the  $k$ -th Chern class of the Hodge bundle of  $\overline{\mathcal{M}}_{g,d}$ . From this Goulden *et al.* deduced the formula

$$\mathcal{H}_g(z) = \sum_{n \geq 0} \frac{1}{n!} \frac{T^n}{(1-T)^{2g-2+n}} \sum_{\substack{0 \leq k \leq g, k_1, \dots, k_n \geq 2 \\ k + \sum k_i = 3g - 3 + n}} (-1)^k \int_{\overline{\mathcal{M}}_{g,n}} \lambda_k \psi_1^{k_1} \cdots \psi_n^{k_n} \quad (40)$$

for the generating series  $\mathcal{H}_g$ . Since the first sum obviously terminates at  $n = 3g - 3$  (because now each  $k_i$  is at least 2), this clearly has the form (34), with top coefficient given by

$$\kappa_{g,5g-5} = \frac{1}{(3g-3)!} \int_{\overline{\mathcal{M}}_{g,3g-3}} \psi_1^2 \cdots \psi_{3g-3}^2,$$

an intersection number that was studied in [32] and shown to have a generating function satisfying a Painlevé differential equation. Similarly, if we look at the asymptotics of (40) for  $T \rightarrow \infty$  instead of  $T \rightarrow 1$ , and compare with (35), we deduce as a corollary the following Hodge integral formula, which seems to be new:

**Corollary.** For all  $g \geq 2$ , we have

$$\sum_{n=0}^{3g-3} \frac{(-1)^n}{n!} \sum_{\substack{0 \leq k \leq g, k_1, \dots, k_n \geq 2 \\ k + \sum k_i = 3g - 3 + n}} (-1)^k \int_{\mathcal{M}_{g,n}} \lambda_k \psi_1^{k_1} \dots \psi_n^{k_n} = \frac{B_{2g}}{2g(2g-2)}.$$

**1.4. Lambert Space.** Consider the vector space  $\mathbb{Q}^{\mathbb{N}}$  of sequences  $(f(1), f(2), \dots)$  of rational numbers, equipped with the convolution ring structure

$$(f * g)(d) = \sum_{e=1}^{d-1} \binom{d}{e} f(e) g(d-e) \quad (f, g \in \mathbb{Q}^{\mathbb{N}}, d \in \mathbb{N}) \tag{41}$$

and with the automorphism  $D$  defined by  $(Df)(d) = df(d)$ , which is a derivation with respect to  $*$ . The ring  $(\mathbb{Q}^{\mathbb{N}}, *)$  is isomorphic to the ring  $z\mathbb{Q}[[z]]$  of formal power series without constant term in an indeterminate  $z$  via the exponential generating series:

$$\mathbb{Q}^{\mathbb{N}} \ni f \mapsto \bar{f} = \bar{f}(z) = \sum_{d \geq 1} f(d) \frac{z^d}{d!} \in z\mathbb{Q}[[z]]. \tag{42}$$

Under this isomorphism  $D$  corresponds to the derivation  $z \frac{d}{dz}$  used before. The space  $\mathbb{Q}^{\mathbb{N}}$  contains the subspace  $\mathcal{A} = (\mathbb{Z}^{\mathbb{N}}) \otimes \mathbb{Q}$  of sequences with bounded denominator, which is closed under multiplication and under the action of  $D$ , but not under that of  $D^{-1}$ . Of course both  $\mathbb{Q}^{\mathbb{N}}$  and  $\mathcal{A}$  are modules of infinite rank over  $\mathbb{Q}[D]$ . We will be interested in a particular subspace  $\Lambda \subset \mathcal{A}$  that we will call the ‘‘Lambert space’’ and that is a free  $\mathbb{Q}[D]$ -module of rank 2. We can define it somewhat artificially by

$$\Lambda = \mathbb{Q}[D]\alpha \oplus \mathbb{Q}[D]\beta, \tag{43}$$

where  $\alpha, \beta \in \mathbb{Q}^{\mathbb{N}}$  are the sequences defined by

$$\alpha - \alpha * \alpha_{-1} = \alpha_{-1}, \quad \beta - \beta * \alpha_{-1} = \alpha \quad (\text{or } \beta * \alpha_{-1} = \alpha * \alpha), \tag{44}$$

where  $\alpha_n$  and  $\beta_n$  for  $n \in \mathbb{Z}$  denote  $D^n \alpha$  and  $D^n \beta$ , respectively, so that  $\Lambda$  has the  $\mathbb{Q}$ -basis  $(\alpha_n, \beta_n)_{n \geq 0}$ . We can solve the equations (44) recursively to compute the first values:

$d$	1	2	3	4	5	6	7	8	...
$\alpha_{-1}(d)$	1	2	9	64	625	7776	117649	2097152	...
$\beta_{-1}(d)$	1	3	17	142	1569	21576	355081	6805296	...

(here we have tabulated  $\alpha_{-1}$  and  $\beta_{-1}$  instead of  $\alpha$  and  $\beta$  because they are already integer-valued). These functions can be given in closed form as

$$\alpha_{-1}(d) = \frac{\alpha(d)}{d} = d^{d-1}, \quad \beta_{-1}(d) = \frac{\beta(d)}{d} = (d-1)! \sum_{r=0}^{d-1} \frac{d^r}{r!} \quad (d \geq 1). \tag{45}$$

In particular,  $\alpha_n$  belongs to  $\mathcal{A}$  for all  $n \in \mathbb{Z}$ , whereas  $\beta_n$  belongs to  $\mathcal{A}$  only for  $n \geq -1$ . (The number  $\beta_{-2}(p)$  has denominator  $p$  for every prime number  $p$ .) We define

the extended Lambert space  $\Lambda^+$  by

$$\Lambda^+ = \bigoplus_{n=-\infty}^{\infty} \mathbb{Q}\alpha_n \oplus \bigoplus_{n=-1}^{\infty} \mathbb{Q}\beta_n \subset \mathcal{A}$$

and define two further subspaces of  $\Lambda^+$  by

$$\Lambda_1 = \bigoplus_{n=-\infty}^{-1} \mathbb{Q}\alpha_n, \quad \Lambda_2 = \bigoplus_{n=-\infty}^{\infty} \mathbb{Q}\alpha_n \oplus \bigoplus_{n=0}^{\infty} \mathbb{Q}\beta_n,$$

so that  $\Lambda_2$  has codimension 1 in  $\Lambda^+$  and each of  $\Lambda$ ,  $\Lambda^+$ , and  $\Lambda_2$  is a  $\mathbb{Q}[D]$ -module.

**Proposition 2.** *Each of the spaces  $\Lambda$ ,  $\Lambda_1$  and  $\Lambda_2$  is a ring with respect to the multiplication (41).*

*Proof.* Set  $T = \overline{\alpha_{-1}} \in z\mathbb{Q}[[z]]$ , where  $\overline{\alpha_{-1}}$  is defined by (42). Then the first of equations (44) says that  $(1 - T)D(T) = T$  or  $\frac{1}{z} dz = \frac{1-T}{T} dT$ , which integrates immediately to (27). Thus the series  $\overline{\alpha_{-1}}$  is given by (28), the Lambert power series (whence the name ‘‘Lambert space’’) and  $\alpha(d)$  is given by the first formula in (45). Similarly, the second or third of equations (44) give  $(1 - T)\overline{\beta} = T/(1 - T)$  and  $T\overline{\beta} = T^2/(1 - T)^2$ , respectively, so  $\overline{\beta} = T/(1 - T)^2$ . (One can derive from this generating function the second formula in (45), but this formula will not be used and we omit the proof.) It follows immediately that  $\overline{\alpha_n}$  and  $\overline{\beta_n}$  for  $n \geq 0$  are polynomials in  $\frac{T}{1-T}$  of degree  $2n + 1$  and  $2n + 2$ , respectively, and that  $\overline{\alpha_{-n}}$  for  $n > 0$  is a polynomial in  $T$  of degree  $T$ , the first values being

$n$	0	1	2	3	4
$\overline{\alpha_n}$	$\frac{T}{1-T}$	$\frac{T}{(1-T)^3}$	$\frac{T+2T^2}{(1-T)^5}$	$\frac{T+8T^2+6T^3}{(1-T)^7}$	$\frac{T+22T^2+58T^3+24T^4}{(1-T)^9}$
$\overline{\beta_n}$	$\frac{T}{(1-T)^2}$	$\frac{T+T^2}{(1-T)^4}$	$\frac{T+5T^2+2T^3}{(1-T)^6}$	$\frac{T+15T^2+26T^3+6T^4}{(1-T)^8}$	$\frac{T+37T^2+168T^3+154T^4+24T^5}{(1-T)^{10}}$
$\overline{\alpha_{-n}}$		$T$	$T - \frac{1}{2}T^2$	$T - \frac{3}{4}T^2 + \frac{1}{6}T^3$	$T - \frac{7}{8}T^2 + \frac{11}{36}T^3 - \frac{1}{24}T^4$

In particular, the spaces  $\Lambda$  (resp.  $\Lambda_1$  or  $\Lambda_2$ ) correspond under the isomorphism  $\mathbb{Q}^N \cong z\mathbb{Q}[[z]] \cong T\mathbb{Q}[[T]]$  to the spaces of polynomials in  $\frac{T}{1-T}$  (resp. polynomials or Laurent polynomials in  $1 - T$ ) that vanish at  $T = 0$ , so each of them is closed under multiplication.  $\square$

**Remark 1.** Despite the easy nature of this proof, the assertion of the proposition corresponds to convolution identities for  $\alpha_n$  and  $\beta_n$  that are not at all obvious, like

$$\sum_{e=1}^{d-1} \binom{d}{e} \left(\frac{e}{d}\right)^{e-3} \left(1 - \frac{e}{d}\right)^{d-e-2} = \frac{5d^3 + 21d^2 + 94d + 12}{12} \in \mathbb{Q}[d].$$

**Remark 2.** The fact that elements of  $\Lambda$  belong to the space  $\mathcal{A}$  of sequences with bounded denominator implies an integrality statement for the Hurwitz numbers (and similarly also for the enumerative functions of graphs and other sequences studied in Section 3). Specifically, the number  $h_{g,d}$ , which from its definition as  $H_{g,d}/(2d + 2g - 2)!$  could have a denominator as large as  $(2d + 2g - 2)!$ , in fact has

denominator at most  $N_g d!$ , for some integer  $N_g$  depending only on  $g$ . On the other hand, we do not know the combinatorial meaning of the divisibility of an element of  $\Lambda$  by a power of  $(1 - T)^{-1}$  (like the divisibility of  $\mathcal{H}_g$  by  $(1 - T)^{-2g+2}$ ), let alone the meaning of the bottom coefficient.

We end this subsection with a few remarks on the asymptotics of sequences in the Lambert space. The two basic sequences  $\alpha$  and  $\beta$  have asymptotic growth given by

$$\begin{aligned} \frac{\alpha(d) e^d}{d!} &= \frac{1}{\sqrt{2\pi d}} \left( 1 - \frac{1}{12d} + \frac{1}{288d^2} + \dots \right), \\ \frac{\beta(d) e^d}{d!} &= \frac{1}{2} - \frac{1}{\sqrt{2\pi d}} \left( \frac{1}{3} + \frac{1}{540d} - \frac{25}{6048d^2} + \dots \right), \end{aligned}$$

and the asymptotic growth of  $\alpha_n(d)$  and  $\beta_n(d)$  for any  $n \in \mathbb{Z}$  is obtained by multiplying these by  $d^n$ , so that we can give the asymptotics of any element  $f \in \Lambda$  to arbitrary order in  $d^{-1/2}$ . We can also see the same thing analytically using the isomorphism (42), by considering the behavior of  $\bar{f}(z)$  as  $z \rightarrow e^{-1}$ . From the proposition and its proof we know that  $\bar{f}(z)$  is a polynomial in  $1/(1 - T)$ , so as  $\varepsilon = 1 - T$  tends to 0 (from above) we have  $\bar{f}(z) \sim C\varepsilon^{-D}$  for some positive integer  $D$  and non-zero rational number  $C$ . (These numbers  $D$  and  $C$  are what we called the “order of the most singular term” and the “top coefficient” in the discussion at the end of the introduction.) But from (27) we get  $1 - ez = 1 - (1 - \varepsilon)e^\varepsilon = \frac{1}{2}\varepsilon^2 + O(\varepsilon^3)$ , so  $\bar{f}(z) \sim 2^{-D/2}C(1 - ez)^{-D/2}$  as  $z$  tends to  $e^{-1}$  from below, and from this and the binomial theorem one gets the asymptotics  $\frac{f(d)}{d!} \sim \frac{2^{-D/2}C}{\Gamma(1+D/2)} d^{D/2-1} e^d$  or  $f(d) \sim \frac{2^{-D/2}C\sqrt{2\pi}}{\Gamma(1+D/2)} d^{d+(D-1)/2}$ , as well as an expansion to higher orders if one wishes.

**1.5. Asymptotics of the Hurwitz numbers.** We end Section 1 with a discussion of the asymptotic behavior of the numbers  $H_{g,d}$  (or  $h_{g,d}$ ) for large values of  $g$  and  $d$ .

The large  $g$  asymptotics for  $H_{g,d}$  with fixed  $d \geq 2$  follow immediately from the corollary to Theorem 1 (Hurwitz’s theorem) and are given by (3). Notice that this asymptotic formula is extremely precise, since the error term is of the order of  $\binom{d-1}{2}^{2g}$  and is hence exponentially smaller than the main term for  $g$  large. A related comment is that the generating function  $C_d(x)$  for  $x \in \mathbb{R}_{>0}$  fixed is also given to high accuracy by the leading term approximation

$$C_d(x) \approx \frac{1}{d!^2} \exp\left(\binom{d}{2}x\right) \quad (d \rightarrow \infty). \tag{46}$$

This is true, although not quite obvious, even for  $x$  small; e.g., for  $x = 0.1$  the ratio of the left- and right-hand sides of (46) differs from 1 by about  $5 \times 10^{-18}$  for  $d = 500$  and  $6 \times 10^{-82}$  for  $d = 2000$ . In particular, the generating series  $\mathcal{H}(x, y)$  defined in (1) is rapidly divergent for any positive values of  $x$  and  $y$ , so that it can only be considered as a formal power series.

In the opposite direction for  $g$  fixed and  $d$  large, the discussion at the end of Section 1.4 together with the formula given in (35) for the “top coefficient” of the

element  $d \mapsto d! h_{g,d}$  of the Lambert space immediately imply the asymptotic statements (5) and (6). It is perhaps worth mentioning that one can also proceed in a somewhat different order, using the differential recursion (25) inductively to derive the formula

$$\mathcal{H}_g(z) \sim \frac{2^{-5(g-1)/2} 24^{-g} c_g}{(5g-1)(5g-3)} (1 - ez)^{-5(g-1)/2} \quad \text{for } z \rightarrow e^{-1}$$

with  $c_g$  satisfying (7) for the asymptotics of each function  $\mathcal{H}_g(z)$  as  $z$  tends to  $e^{-1}$  from below and deducing the asymptotic formula for its Taylor coefficients from this. We suppress the details, since they are a little lengthy and basically just reproduce the arguments used in the proof of Theorem 3. Note that each individual term of the expansion  $\mathcal{H}(x, y)$  as  $\sum_g \mathcal{H}_g(x^2 y) x^{2g-2}$  converges for  $|x^2 y| < 1/e$ , but that the whole series does not converge for any positive values of  $x$  and  $y$ , as already mentioned above. We can also see why this should be so from the “fixed  $g$ ” asymptotics with  $g$  large. The numbers  $c_g$  defined by (7) or (8) grow more than exponentially, like

$$c_g \sim \frac{\sqrt{3/5}}{\pi} \left( \frac{5\sqrt{2}}{e} g \right)^{2g} \left( \frac{1}{g} + \frac{1}{6g^2} + \frac{1}{72g^3} - \frac{14459}{810000g^4} + \dots \right)$$

as  $g \rightarrow \infty$ . (To see that this should be true up to a constant independent of  $g$ , one notes that the recursion (7) has a formal solution  $C \cdot 50^g (g-1)!^2 P_g$  for a unique power series  $P_g = 1 - \frac{49}{3g^3} + \dots$ . The rigorous proof that this formal solution gives the correct asymptotics of  $c_g$ , and the determination of the constant  $C = \frac{\sqrt{3/5}}{2\pi^2}$ , are given in various places in the literature [34], [33], [25], [45].) The renormalized values occurring in (6) decay more than exponentially, like  $(e/720g)^{g/2}$ , but this is not enough to ensure convergence of the double series since together with (6) (assuming that this estimate holds uniformly when  $g$  and  $d$  are both large, which we do not know) this would say that the  $(g, d)$  term of  $\mathcal{H}$  looks roughly like  $(ed^5 x^4 / 720g)^{g/2} (ex^2 y)^d$  and hence grows more than exponentially rapidly in the range when  $d$  goes to infinity more rapidly than  $g^{1/5}$ . It remains an interesting open question to find a uniform bound for  $h_{g,d}$  or for  $H_{g,d}$  in terms of elementary functions.

**Remark.** Another interesting question about enumerations of Hurwitz covers of a torus has been systematically studied in [8], where the Lambert ring should be replaced by the ring of quasi-modular forms. We hope to continue the study of Hurwitz covers of a Riemann surface of higher genus in future.

## 2. COMPUTING $H_{g,d}$ FROM THE $\mathbb{P}^1$ FROBENIUS MANIFOLD

In this section we describe a completely different approach, and a completely different proof of the formula (34), using the Gromov–Witten (GW) invariants of  $\mathbb{P}^1$ . Unlike the preceding section, this one is not elementary and we will assume familiarity with the theory of Frobenius manifolds.

The GW invariants of  $\mathbb{P}^1$  are encoded by the generating series (“free energy”)

$$\mathcal{F}^{\mathbb{P}^1} = \sum_{g \geq 0} \mathcal{F}_g^{\mathbb{P}^1} \epsilon^{2g-2},$$

where the genus  $g$  part  $\mathcal{F}_g^{\mathbb{P}^1}$  is given by

$$\mathcal{F}_g^{\mathbb{P}^1} = \sum_{n, d \geq 0} \frac{1}{n!} \sum_{\substack{p_1, \dots, p_n \geq 0 \\ 1 \leq \alpha_1, \dots, \alpha_n \leq 2}} t_{p_1}^{\alpha_1} \dots t_{p_n}^{\alpha_n} \int_{[\overline{\mathcal{M}}_{g,n}(\mathbb{P}^1, d)]^{\text{virt}}} \text{ev}_1^*(\phi_{\alpha_1}) \dots \text{ev}_n^*(\phi_{\alpha_n}) \psi_1^{p_1} \dots \psi_n^{p_n}.$$

Here  $t_p^1, t_p^2$  ( $p \geq 0$ ) are indeterminates,  $\phi_1 = 1$  and  $\phi_2 = \omega$  are the standard basis of  $H^*(\mathbb{P}^1)$  (with  $\omega$  normalized by  $\int_{\mathbb{P}^1} \omega = 1$ ),  $\overline{\mathcal{M}}_{g,n}(\mathbb{P}^1, d)$  is the moduli space of stable maps of degree  $d$  of curves of genus  $g$  with  $n$  marked points to the target  $\mathbb{P}^1$ ,  $\text{ev}_i$  denotes the  $i$ -th evaluation map and  $\psi_i$  the first Chern class of the  $i$ -th tautological line bundle on  $\overline{\mathcal{M}}_{g,n}(\mathbb{P}^1, d)$ . It was proved by Pandharipande [43] that the Hurwitz numbers  $H_{g,d}$  coincide with the following special GW invariants of  $\mathbb{P}^1$

$$H_{g,d} = \int_{[\overline{\mathcal{M}}_{g,n}(\mathbb{P}^1, d)]^{\text{virt}}} \text{ev}_1^*(\omega) \dots \text{ev}_n^*(\omega) \psi_1 \dots \psi_n,$$

where  $n$  is defined as  $2g + 2d - 2$  (the integral vanishes for other values of  $n$ ), so  $\mathcal{F}_g^{\mathbb{P}^1}$  is related to the generating series  $\mathcal{H}_g$  of Section 1 by

$$t^{2g-2} \mathcal{H}_g(t^2) = \mathcal{F}_g^{\mathbb{P}^1} \Big|_{t_1^2=t, t_p^\alpha=0 \text{ otherwise}}, \tag{47}$$

where the variable  $t^2$  corresponds to the variable  $z = x^2y$  of Section 1.3 and the full generating series  $\mathcal{F}^{\mathbb{P}^1}$  coincides with our  $\mathcal{H}(x, y)$  under the substitution  $t = \sqrt{z}$ ,  $\epsilon = x/\sqrt{z} = 1/y$ .

We now apply the DZ approach to calculating these special GW invariants. Even though only the specialization  $t_p^\alpha = \delta_{p_1} \delta_{\alpha 2} t$  of the variables  $t_p^\alpha$  is used in (47), we have to work with the 2-variable specialization with a second non-zero variable  $t_0^1 = s$ , and will only restrict to  $s = 0$  at the end. The potential of the  $\mathbb{P}^1$  Frobenius manifold is given by

$$F = \frac{uv^2}{2} + e^u,$$

where  $u, v$  are the flat coordinates ( $\partial_v$  gives the unit vector field), which in turn will be expressed as power series in  $t$  and  $s$ . The power series  $(u, v)$  is the so-called topological solution, obtained by solving the following genus zero Euler–Lagrange equation (see details in [9], [18]):

$$\frac{\partial \Phi}{\partial u} = \frac{\partial \Phi}{\partial v} = 0, \quad \text{where} \quad \Phi = su + t \frac{\partial F}{\partial u} - uv = su + t \left( \frac{v^2}{2} + e^u \right) - uv.$$

This has a unique solution  $u(s, t), v(s, t) \in \mathbb{C}[[s, t]]$  satisfying  $u(s, 0) = 0, v(s, 0) = s$ , given by

$$u = st + T, \quad v = s + T/t \quad \text{with } T \text{ defined by } t^2 e^{st} = T e^{-T}. \tag{48}$$

The first several terms of the expansions of  $u, v$  near  $s = 0, t = 0$  read

$$u = st + t^2 + st^3 + \frac{s+2}{2}t^4 + \frac{s^3+12s}{6}t^5 + \frac{s^4+48s^2+36}{24}t^6 + \dots,$$

$$v = s + t + st^2 + \frac{s^2+2}{2}t^3 + \frac{s^3+12s}{6}t^4 + \frac{s^4+48s^2+36}{24}t^5 + \dots.$$

Let us now compute  $\mathcal{F}_g^{\mathbb{P}^1}$ . Begin with  $g = 0$ . By [9], [19], we have

$$\mathcal{F}_0^{\mathbb{P}^1} = \frac{s^2}{2} \Omega_{1,0;1,0} - s \Omega_{1,0;1,1} + st \Omega_{1,0;2,1} + \frac{1}{2} \Omega_{1,1;1,1} - t \Omega_{1,1;2,1} + \frac{t^2}{2} \Omega_{2,1;2,1},$$

where  $\Omega_{\alpha,p;\beta,q}$  are the genus 0 two-point functions. Substituting the known explicit expressions

$$\Omega_{1,1;1,1} = (2 - 2u + u^2)e^u + uv^2, \quad \Omega_{1,1;2,1} = uv e^u + \frac{v^3}{3}, \quad \Omega_{2,1;2,1} = \frac{e^{2u}}{2} + v^2 e^u$$

into this formula and specializing to  $s = 0$ , we find

$$\mathcal{F}_0^{\mathbb{P}^1} \Big|_{s=0} = \frac{2T^3 - 9T^2 + 12T}{12t^2},$$

which in view of (47) agrees with our previous formula (30) (with  $Te^{-T} = t^2 = z$ ).

We next consider  $g = 1$ . The generating function  $\mathcal{F}_g^{\mathbb{P}^1}$  in this case has the form

$$\mathcal{F}_1^{\mathbb{P}^1} = \frac{1}{24} \log(v_s^2 - w u_s^2) - \frac{u}{24}.$$

Substituting the solution (48) into this formula and using (47) we obtain

$$\mathcal{H}_1(z) = \frac{1}{24} \log \frac{1}{1-T} - \frac{T}{24},$$

again in agreement with our previous formula (32).

The expression for the genus two part of the free energy is more involved [18], [11]:

$$\begin{aligned} 5760 \mathcal{F}_2^{\mathbb{P}^1} = & -\frac{w^2}{\Delta^4} [512u_s^3 v_s v_{ss}^3 + 384w u_s^3 v_{ss} (u_s^2 + 2u_{ss}) (u_s^2 v_s + 2u_{ss} v_s - 2u_s v_{ss}) \\ & - 64w^2 u_s^4 (u_s^2 + 2u_{ss})^3] \\ & - \frac{w}{\Delta^3} [256u_s v_s v_{ss}^3 + 12w u_s (28u_s^4 v_s v_{ss} + 116u_s^2 u_{ss} v_s v_{ss} + 64u_s^2 v_s v_{ss} \\ & + 28u_s v_s u_{sss} v_{ss} - 69u_s^3 v_{ss}^2 - 128u_s u_{ss} v_{ss}^2 + 14u_s^3 v_s v_{sss} + 28u_s v_s u_{ss} v_{sss} \\ & - 28u_s^2 v_{ss} v_{sss}) - w^2 u_s^2 (u_s^2 + 2u_{ss}) (121u_s^4 + 538u_s^2 u_{ss} + 256u_{ss}^2 + 168u_s u_{sss})] \\ & + \frac{w}{\Delta^2} [-2(42u_s^3 v_s v_{ss} + 126u_s u_{ss} v_s v_{ss} + 42u_{sss} v_s v_{ss} - 95u_s^2 v_{ss}^2 - 96u_{ss} v_{ss}^2 \\ & + 30u_s^2 v_s v_{sss} + 42u_{ss} v_s v_{sss} - 126u_s v_{ss} v_{sss} + 20u_s v_s v_{ssss}) \\ & + w(72u_s^6 + 479u_s^4 u_{ss} + 626u_s^2 u_{ss}^2 + 64u_{ss}^3 \\ & + 224u_s^3 u_{sss} + 252u_s u_{ss} u_{sss} + 40u_s^2 u_{ssss})] \\ & + \frac{1}{\Delta} [22v_{ss}^2 - 24v_s v_{sss} + w(17u_s^4 + 102u_s^2 u_{ss} + 56u_{ss}^2 + 68u_s u_{sss} + 20u_{ssss})] + 7u_{ss}, \end{aligned} \tag{49}$$

where  $\Delta = v_s^2 - w u_s^2$ . Using the expressions (48) for  $u, v$  we find that  $\Delta|_{s=0} = \frac{1}{1-T}$ , so we obtain

$$\mathcal{H}_2(z) = \frac{1}{1440} \frac{T^2(1+6T)}{(1-T)^5},$$

in agreement with the first formula in (33).

To proceed to higher genera, one can solve, recursively in  $g$ , the loop equation [18], [19]. Noting that

$$\begin{aligned} v(0, t) &= \frac{T}{t}, & u(0, t) &= T, & v_s(0, t) &= \frac{1}{1-T}, & u_s(0, t) &= \frac{t}{1-T}, \\ \frac{\partial^\ell v}{\partial s^\ell}(0, t) &= t^{\ell-1} D^\ell(T), & \frac{\partial^\ell u}{\partial s^\ell}(0, z) &= t^\ell D^\ell(T) \quad (\ell \geq 2), \end{aligned}$$

we find by the  $(3g - 3)$ -Lemma [18] that the function  $\mathcal{H}_g$  has the form (34) for all values of  $g$  (Goulden–Jackson–Vakil theorem). As the above solution for  $g = 2$  makes clear, it would not be easy to calculate  $\mathcal{H}_g$  explicitly this way for large values of  $g$ . On the other hand, it should be emphasized that the formulas  $\mathcal{F}_g^{\mathbb{P}^1}$  obtained by this approach, such as the formula (49) for the case  $g = 2$ , would hold without any modification for the general GW invariants (i.e., without the specialization to  $t_p^\alpha = 0$  for  $(p, \alpha) \neq (1, 2)$ ), with only the form of  $\Phi$  changing. It is worth noting that the DZ loop equation for solving  $\mathcal{F}_g^{\mathbb{P}^1}$  is quadratic, and also universal in all semisimple homogeneous cohomological field theories, with the same form in all cases and the dependence on the Frobenius manifold arising only through its Euler vector field and periods.

### 3. CONNECTION TO VARIOUS MODELS IN ENUMERATIVE GEOMETRY

In this section we will look in detail at three other problems whose solutions belong to the Lambert space defined in Section 1.4: the enumerations of ordinary graphs and of ribbon graphs, and the computation of certain Hodge integrals that include the Hurwitz numbers as a special case. The discussion of the first case is elementary and self-contained, but the discussion of the two other cases assumes some familiarity with moduli space theory.

**3.1. Enumeration of graphs.** In this subsection we discuss the enumeration of graphs. Denote by  $G_{g,d}$  the number of connected graphs with  $d$  vertices and  $g$  independent loops (i.e., first Betti number  $g$ ), where graphs are allowed to have multiple (non-oriented) edges and by “number” we mean the weighted number in which each graph  $\Gamma$  is counted with multiplicity  $1/|\text{Aut}(\Gamma)|$ , the reciprocal of its number of symmetries. For instance,  $G_{1,2} = \frac{3}{4}$  because there are two graphs with two vertices and one loop, a “tadpole” with a symmetry group of order 2 (reflect the loop) and a 2-gon with symmetry group of order 4 (interchange the two vertices or the two edges), and  $G_{g,1} = 1/2^g g!$  for all  $g$  because the only 1-vertex graph with  $g$  loops is a bouquet of  $g$  circles that can be arbitrarily re-oriented or permuted. In this subsection we describe the numbers  $G_{g,d}$  when either  $d$  or  $g$  is fixed, finding in both cases results exactly analogous to those for the numbers  $h_{g,d}$ . Most or all of

these results are certainly known (see for example [46]), but we give a presentation that is as close as possible to the one for the Hurwitz case.

We begin with a discussion of the case of fixed  $d$ .

**Theorem 4.** *For each  $d \geq 1$ , the generating series*

$$B_d(x) = \sum_{g \geq 0} G_{g,d} x^{g+d-1}$$

*of graphs with  $d$  vertices is a polynomial of degree  $d^2$  in  $e^{x/2}$  of the form*

$$B_d(x) = \frac{1}{d!} \sum_{\substack{d \leq m \leq d^2 \\ m \equiv d \pmod{2}}} c_{d,m} e^{mx/2}, \quad (50)$$

*where the coefficients  $c_{d,m}$  are integers with  $c_{d,d^2} = 1$ ,  $c_{d,m} = 0$  for  $d^2 - 2d + 2 < m < d^2$ , and  $c_{d,d} = (-1)^{d-1}(d-1)!$ . These polynomials are determined completely either by the linear recursion*

$$B_d(x) = \frac{e^{d^2 x/2}}{d!} - \frac{1}{d} \sum_{k=1}^{d-1} \frac{d-k}{k!} e^{k^2 x/2} B_{d-k}(x) \quad (51)$$

*or by the quadratic recursion*

$$B_d(x) = \frac{1}{d^2 - d} \sum_{k=1}^{d-1} k(d-k) (e^{kx} - 1) B_k(x) B_{d-k}(x) \quad (d \geq 2) \quad (52)$$

*or by the quadratic differential recursion*

$$\left(2 \frac{d}{dx} - d^2\right) B_d(x) = \sum_{k=1}^{d-1} k(d-k) B_k(x) B_{d-k}(x). \quad (53)$$

**Corollary.** *The numbers  $G_{g,d}$  for fixed  $d$  have the form*

$$G_{g,d} = \frac{1}{d! (d+g-1)!} \sum_{\substack{1 \leq m \leq d^2 \\ m \equiv d \pmod{2}}} c_{d,m} (m/2)^{g+d-1}$$

*with  $c_{d,m}$  as above, the first values being given by*

$d$	$2^w w! d! G_{g,d}$ , where $w = g + d - 1$
1	1
2	$4^w - 2^w$
3	$9^w - 3 \cdot 5^w + 2 \cdot 3^w$
4	$16^w - 4 \cdot 10^w - 3 \cdot 8^w + 12 \cdot 6^w - 6 \cdot 4^w$
5	$25^w - 5 \cdot 17^w - 10 \cdot 13^w + 20 \cdot 11^w + 30 \cdot 9^w - 60 \cdot 7^w + 24 \cdot 5^w$

*Proof.* The proof follows the same idea as the second approach to Theorem 1 in Section 1.2, via disconnected Hurwitz numbers. For  $k, d \geq 0$  let  $N_{k,d}$  denote the number of graphs (not necessarily connected) with  $d$  vertices and  $k$  edges, where “number” has the same meaning as for  $G_{g,d}$  above, and let

$$Z_G = Z_G(x, y) = \sum_{\Gamma} \frac{x^{e(\Gamma)} y^{v(\Gamma)}}{|\text{Aut}(\Gamma)|} = \sum_{k, d \geq 0} N_{k,d} x^k y^d \tag{54}$$

(where  $e(\Gamma)$  and  $v(\Gamma)$  denote the number of edges and number of vertices of a graph  $\mathcal{G}$ ) be the corresponding generating series, called the partition function of graph counting. By the usual principle that the generating series for all objects of a given type is the exponential of the generating series for the connected ones, together with the observation that a connected graph with  $d$  vertices and  $g$  loops has Euler number  $1 - g$  and hence has  $g + d - 1$  edges, we have

$$Z_G = Z_G(x, y) = e^{\mathcal{G}(x,y)}, \tag{55}$$

where

$$\mathcal{G}(x, y) = \sum_{\Gamma \text{ connected}} \frac{x^{e(\Gamma)} y^{v(\Gamma)}}{|\text{Aut}(\Gamma)|} = \sum_{g \geq 0, d \geq 1} G_{g,d} x^{g+d-1} y^d = \sum_{d=1}^{\infty} B_d(x) y^d \tag{56}$$

is the two-variable generating series for connected graphs. On the other hand, it is easy to see that  $N_{k,d} = d^{2k} / 2^k k! d!$ , since there are  $2^k k! d!$  ways to number and orient the edges and number the vertices (after which there are no longer any automorphism) and then  $d^{2k}$  ways to map the  $2k$  numbered endpoints of the edges to the  $d$  numbered vertices. This gives the closed formula

$$Z_G(x, y) = \sum_{d \geq 0} \frac{y^d}{d!} e^{d^2 x / 2}. \tag{57}$$

for the partition function  $Z_G$ . Formulas (55)–(57) already imply that the coefficients  $B_d(x)$  of  $\mathcal{G}(x, y)$  belong to  $\mathbb{Q}[e^{x/2}]$  and have the form given in (50). Formula (57) also implies the differential equations

$$DZ_G(x, y) = y e^{x/2} Z_G(x, ye^x), \quad D^2 Z_G(x, y) = 2 \frac{\partial Z_G(x, y)}{\partial x} \tag{58}$$

for  $Z_G$ , where  $D = y \frac{\partial}{\partial y}$ , and these and (55) give the differential equations

$$D(Z_G) = D(\mathcal{G}) Z_G, \tag{59}$$

$$(D^2 - D)\mathcal{G}(x, y) = D\mathcal{G}(x, y)(D\mathcal{G}(x, ye^x) - D\mathcal{G}(x, y)), \tag{60}$$

$$(D\mathcal{G})^2 + D^2\mathcal{G} = 2 \frac{\partial \mathcal{G}}{\partial x} \tag{61}$$

for  $\mathcal{G}$ , which in turn are equivalent to the three recursions (51)–(53). □

We now turn to the opposite case, when  $g$  is fixed.

**Theorem 5.** *The generating series*

$$\mathcal{G}_g(z) = \sum_{d=1}^{\infty} G_{g,d} z^d \tag{62}$$

belongs to the extended Lambert space for all  $g$ . It is given for  $g \leq 2$  by

$$\mathcal{G}_0(z) = T - \frac{T^2}{2}, \quad \mathcal{G}_1(z) = \frac{1}{2} \log \frac{1}{1-T}, \quad \mathcal{G}_2(z) = \frac{3T + 2T^2}{24(1-T)^3} \quad (63)$$

(where  $z = T e^{-T}$  as usual) and in general by an expression of the form

$$\mathcal{G}_g = \sum_{i=g-1}^{3g-3} \frac{\lambda_{g,i}}{(1-T)^i} \quad (64)$$

for all  $g \geq 2$ , with top and bottom coefficients given by

$$\lambda_{g,g-1} = \frac{B_g}{g(g-1)}, \quad \lambda_{g,3g-3} = b_{g-1}, \quad (65)$$

where  $B_n$  denotes the  $n$ th Bernoulli number and the numbers  $b_r$  are defined by the generating series

$$\begin{aligned} \sum_{r=1}^{\infty} b_r x^r &= \log \left( \sum_{m=0}^{\infty} \frac{(6m)!}{(3m)!(2m)!} \left( \frac{x}{288} \right)^m \right) \\ &= \frac{5}{24} x + \frac{5}{16} x^2 + \frac{1105}{1152} x^3 + \frac{565}{128} x^4 + \frac{82825}{3072} x^5 + \dots \end{aligned} \quad (66)$$

*Proof.* We will outline three proofs. We first give a “pure thought” proof of the statements that  $\mathcal{G}_g$  is given by (63) for  $g \leq 1$  and is a polynomial in  $1/(1-T)$  of degree  $3g-3$  and leading coefficient  $b_{g-1}$  for all  $g \geq 2$ . We then sketch two other proofs of these statements, and also of the formula for the “bottom” coefficient in (65), based on the closed formula (57) for  $e^{\mathcal{G}}$ .

Consider first the case  $g = 0$ . The number  $G_{0,d}$  counts trees (connected and simply connected graphs) on  $d$  vertices. The number  $dG_{0,d}$  counts rooted trees (trees with one marked vertex, the “root”), because there are  $d$  ways to choose the root of a given tree and the factors  $1/|\text{Aut}(\Gamma)|$  in the definitions of  $G_{0,d}$  and  $G_{1,d}$  ensure that everything works correctly. Denote by  $T(z) = D\mathcal{G}_0(z) = \sum_{d \geq 1} dG_{0,d} z^d$  the generating series of rooted trees. Then the generating series of rooted trees in which the root has valency  $k$  is clearly  $zT(z)^k/k!$ , with the factor  $z$  corresponding to the root and the factor  $T(z)^k/k!$  to the fact that the complement of the root is the union of  $k$  rooted trees (the root being the other end of one of the edges going to the original root). Since the root of every tree has some valency, we get  $T(z) = \sum_{k=0}^{\infty} zT(z)^k/k! = ze^{T(z)}$ , and hence  $T(z)$  is indeed equal to the series defined by (27). This proves the case  $g = 0$  of (63), and the Lagrange inversion formula (28) gives the explicit formula  $d!G_{0,d} = d^{d-2}$ , which is Cayley’s famous result for the number of trees on  $d$  given vertices.

Next consider  $g = 1$ . A graph with Betti number 1 has a unique cycle (with no backtracking). If this cycle contains  $k$  vertices, then the graph is the union of the cycle and  $k$  rooted trees, the roots being the vertices on the cycle, so the contribution to  $G_1(z)$  coming from all such graphs is  $T(z)^k/2k$ , with the factor  $2k$  corresponding to the automorphisms of the cycle (rotations and reflection). This gives  $\mathcal{G}_1(z) = \sum_{k=1}^{\infty} T(z)^k/2k$  and hence proves (63) for  $g = 1$ .

Now suppose that  $g \geq 2$ . There is a map from the set  $\text{Gr}_g$  of graphs with  $g$  loops to the set  $\text{Gr}_g^{\geq 3}$  of graphs with  $g$  loops and having only vertices of valency  $\geq 3$ : if a graph has a 1-valent vertices, we remove this vertex and the corresponding edge and keep doing this until there are no 1-valent vertices left, and if the graph then has any 2-valent vertices, we remove them and fuse the corresponding two edges into a single new one. This map is clearly surjective, with the graphs that map to a given ( $\geq 3$ )-valent graph  $\Gamma$  being given by attaching a rooted tree to any vertex of  $\Gamma$ , which gets identified with the root of the tree, and any number  $n \geq 0$  of rooted trees to any edge of  $\Gamma$ , where the roots become new internal points on that edge. Thus the contribution to the generating function  $\mathcal{G}_g(z)$  of each vertex of  $\Gamma$  is  $T(z)$ , and the contribution of each edge is  $\sum_{n=0}^{\infty} T(z)^n = 1/(1 - T(z))$ , giving the identity

$$\mathcal{G}_g(Te^{-T}) = \sum_{\Gamma \in \text{Gr}_g^{\geq 3}} \frac{1}{|\text{Aut}(\Gamma)|} \frac{T^{v(\Gamma)}}{(1 - T)^{e(\Gamma)}} = \sum_{n=1}^{\infty} G_{g,n}^{(\geq 3)} \frac{T^n}{(1 - T)^{g+n-1}}, \tag{67}$$

where  $G_{g,n}^{(\geq 3)}$  is the weighted number of connected ( $\geq 3$ )-valent graphs with  $n$  vertices and  $g$  loops. On the other hand if such a graph  $\Gamma$  has  $n_i$  vertices of valency  $i$ , then  $n$  and  $k = e(\Gamma) = n + g - 1$  are given by  $n = n_3 + n_4 + \dots$  and  $2k = 3n_3 + 4n_4 + \dots$ , respectively, so  $n_3 + 2n_4 + \dots = 2g - 2$  and in particular  $n \leq 2g - 2$ . This shows that  $\mathcal{G}_g(z)$  for  $g \geq 2$  is a polynomial of degree  $3g - 3$  in  $1/(1 - T)$  with top coefficient equal to the number of connected trivalent graphs with first Betti number  $g$ . But this number is clearly equal to  $b_{g-1}$  with  $b_r$  given by (66), since the weighted number of not-necessarily-connected trivalent graphs with  $d$  vertices and  $k$  edges is clearly non-zero only if  $d = 2m$  and  $k = 3m$  for some  $m$  and is then equal to  $\frac{(6m)!}{288^m (3m)! (2m)!}$ . (Start with a set of  $3m$  labelled and oriented edges and  $2m$  labelled  $Y$ -shaped vertices; then there are  $(6m)!$  ways to identify the  $6m$  half-edges of the two sets and we must divide by the  $2^{3m}(3m)!$  ways to renumber and re-orient the edges and the  $3!^{2m}(2m)!$  ways to renumber and relabel the vertices.) This concludes the proof of Theorem 5, except for the statement about the lowest-order term in (64).

Note the similarity of (67) to (40), with a power series in  $z$  becoming a polynomial in  $(1 - T)^{-1}$  by the same mechanism in both cases: in the ELSV case, the original sum (39) permitted all values  $k_i \geq 0$  and was infinite, while the new sum (40) has  $k_i \geq 2$  and hence is finite because  $\sum (i - 1)k_i = 3g - 3 - k$  is fixed, while in the graph case the original generating function  $\mathcal{G}_g(z)$  includes all valencies  $r \geq 1$  but the sum in (67) has valencies  $r \geq 3$  and hence is bounded because  $\sum (r - 2)n_r = 2g - 2$ . In both cases the reduction is achieved using the so-called string and dilaton equations [32], [28], [46], [44], [13], [15] that arise in these models.

We now briefly indicate the two alternative proofs mentioned above of the statement that the function  $d \mapsto d!G_{g,d}$  belongs to the (extended) Lambert space for  $g \geq 0$ . The first is exactly similar to the proof of the corresponding statement for the Hurwitz numbers given in Section 1.3. We substitute the expansion

$$\mathcal{G}(x, y) = \sum_{g=0}^{\infty} \mathcal{G}_g(xy) x^{g-1}$$

into the recursive functional equation (60) for  $\mathcal{G}$  to get the recursive differential equation

$$\mathcal{G}_g^{(2)} - \mathcal{G}_g^{(1)} = \sum_{\substack{g_1, g_2 \geq 0, \ell \geq 1 \\ g_1 + g_2 + \ell = g + 1}} \frac{1}{\ell!} \mathcal{G}_{g_1}^{(1)} \mathcal{G}_{g_2}^{(\ell+1)} \tag{68}$$

for the generating series  $\mathcal{G}_g$ , and then argue exactly as in the proof of Theorem 3 to get all of the assertions of Theorem 5, including the statement about the “bottom” coefficients that was not included in our first proof. We omit the details. About the bottom coefficients we remark only that it is surprising that the formula is so similar to the one in the Hurwitz case (except for the doubling of the index). About the top coefficients, it follows from (68) that the second of equations (65) holds with  $b_r$  defined inductively by

$$\tilde{b}_1 = 1, \quad \tilde{b}_r = 6r\tilde{b}_{r-1} + 5 \sum_{r_1+r_2=r-1} \tilde{b}_{r_1}\tilde{b}_{r_2} \quad \text{for } r \geq 2, \tag{69}$$

where  $\tilde{b}_r = \frac{6r}{5} 4^r b_r$  (which is an integer by virtue of this recursion). This in fact coincides with the generating function definition (66) of the  $b_r$ , because by logarithmically differentiating (66) and using the linear differential equation satisfied by the hypergeometric function  $\sum \frac{(6m)!}{(2m)!(3m)!} t^m$  one obtains the *Riccati equation*

$$\frac{dV}{dX} - \frac{1}{2}V^2 + \frac{1}{2}X = 0,$$

for the generating function  $V = -X^{1/2} + \frac{1}{2}X^{-1} + 3 \sum_{r \geq 1} r b_r X^{-1-3r/2}$ , and this in turn is equivalent to the recursion (69). We also mention that, as well as their combinatorial interpretation in terms of counting trivalent graphs, the numbers  $b_r$  occur in the Faber–Zagier formulas [22] for the first relation in the tautological subring of the cohomology ring (or Chow ring) of the moduli spaces of curves; see also for example [2], [44], [5] for some related questions where the numbers  $b_r$  appear.

Finally, we can prove Theorem 5 directly from (55) and (57) by an argument using Gaussian integrals, and we sketch this method briefly also since it is simple and works in many situations. Substituting the identity

$$e^{n^2x/2} = \frac{1}{\sqrt{2\pi x}} \int_{-\infty}^{\infty} \exp(nu - u^2/2x) du$$

into (57), we get

$$Z_{\mathcal{G}}(x, z/x) = \frac{1}{\sqrt{2\pi x}} \int_{-\infty}^{\infty} \exp\left(\frac{ze^u - u^2/2}{x}\right) du.$$

Consider the expansion of this for  $x$  small by the method of stationary phase. The argument of the exponential takes on its maximum when its derivative  $(ze^u - u)/x$  vanishes, so at  $u = T$ , where  $T$  is defined by (27). Expanding around this point,

we find

$$\begin{aligned} Z_G(x, z/x) &= \frac{1}{\sqrt{2\pi x}} \int_{-\infty}^{\infty} \exp\left(\frac{ze^{T+v} - (T+v)^2/2}{x}\right) dv \\ &= \frac{1}{\sqrt{2\pi x}} \exp\left(\frac{T - T^2/2}{x}\right) \int_{-\infty}^{\infty} \exp\left(-\frac{1-T}{2x} v^2 + \sum_{r=3}^{\infty} \frac{T}{r! x} v^r\right) dv \\ &= \exp\left(\frac{T - T^2/2}{x} + \frac{1}{2} \log \frac{1}{1-T}\right) \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-w^2/2} \left(1 + \frac{Tw^3}{6(1-T)^{3/2}} \sqrt{x}\right. \\ &\quad \left. + \left(\frac{T^2w^6}{72(1-T)^3} + \frac{Tw^4}{24(1-T)^2}\right) x + \dots\right) dw \\ &= \exp\left(\frac{T - T^2/2}{x} + \frac{1}{2} \log \frac{1}{1-T} + \left(\frac{5T^2}{24(1-T)^3} + \frac{T}{8(1-T)^2}\right) x + \dots\right), \end{aligned}$$

where in the last two lines we have substituted  $v = (\frac{x}{1-T})^{1/2} w$  and used that  $\frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-w^2/2} w^n dw$  equals  $(n-1)!!$  for  $n$  even and 0 for  $n$  odd. This shows that  $\log Z_G(x, z/x)$  has the form  $\sum_{g \geq 0} \mathcal{G}_g(z) x^{g-1}$  with  $\mathcal{G}_0, \mathcal{G}_1$  and  $\mathcal{G}_2$  as in (63) and, by expanding further, that the general coefficient  $\mathcal{G}_g(x)$  is always a polynomial in  $1/(1-T)$ . One also sees that this polynomial has degree  $3g-3$  and top coefficient  $b_{g-1}$  as defined by (66), since the highest power of  $1/(1-T)$  for a given power of  $x$  comes from exponentiating the first term  $Tv^3/6x$  of the argument of the exponential function. □

The following corollary follows immediately by Theorem 5 and by the general remarks made in Section 1.4 about the asymptotics of sequences belonging to the Lambert space.

**Corollary.** *The asymptotic growth of the numbers  $G_{g,d}$  for fixed  $g \geq 0$  is given by*

$$G_{g,d} \sim \frac{5\tilde{b}_{g-1}}{2^{(7g-3)/2} \Gamma(\frac{3g-1}{2})} d^{(3g-5)/2} e^d \quad (d \rightarrow \infty), \tag{70}$$

where  $\tilde{b}_r = \frac{6r}{5} 4^r b_r$  are the integers defined by (69) for  $r \geq 1$  and  $\tilde{b}_{-1} = -\frac{1}{10}$ ,  $\tilde{b}_0 = \frac{1}{5}$ .

Before ending this subsection, we mention that the interesting relationship between enumeration of graphs and Hurwitz numbers discussed here should be revealed in other, more subtle ways. For example, Cayley’s famous formula can be proved by counting the multiplicity of the quasi-homogeneous Lyashko–Looijenga mapping assigning to a polynomial of degree  $d$  with  $d-1$  critical values. For this interesting direction, we omit the details but only refer to Arnold’s paper [1] and the references therein. We hope to return to this question in the future.

**3.2. Ribbon graphs.** We now turn to our second model, related to the counting of ribbon graphs with even valencies. Here the corresponding generating function  $\mathcal{F}$  — called the “GUE (Gaussian unitary ensemble) free energy of even couplings” because of its alternative method of calculation using matrix integrals — and its genus

$g$  parts  $\mathcal{F}_g$  are defined by

$$\begin{aligned} \mathcal{F} &= \mathcal{F}(X, \mathbf{s}; \epsilon) \\ &= \frac{X^2}{2\epsilon^2} \left( \log X - \frac{3}{2} \right) - \frac{1}{12} \log X + \zeta'(-1) + \sum_{g \geq 2} \epsilon^{2g-2} \frac{B_{2g}}{4g(g-1)X^{2g-2}} \\ &\quad + \sum_{g \geq 0} \epsilon^{2g-2} \sum_{n \geq 0} \sum_{j_1, \dots, j_n \geq 1} \left( \sum_{\Gamma} \frac{1}{\#\text{Aut}(\Gamma)} \right) s_{j_1} \dots s_{j_n} X^{2-2g-(n-|j|)}, \\ &=: \sum_{g=0}^{\infty} \epsilon^{2g-2} \mathcal{F}_g(X, \mathbf{s}), \end{aligned} \quad (71)$$

where the internal sum is taken over all connected oriented ribbon graphs of genus  $g$  with labelled half-edges and with unlabelled vertices of valencies  $2j_1, \dots, 2j_n$ ,  $\text{Aut}(\Gamma)$  denotes the symmetry group of  $\Gamma$ , and  $|j| := j_1 + \dots + j_n$ . Actually, the full GUE free energy computes the numbers of ribbon graphs with arbitrary valencies, but here we can treat only the case when all valencies are even. Also, we will actually consider a specialization of (71), which seems to be new. The possibility of effective computation of the new model is due to the recently discovered Hodge–GUE correspondence [17], [14].

It will be convenient to renormalize the couplings by setting  $\bar{s}_k = \binom{2k}{k} s_k$ . We define  $w = w(X, \mathbf{s})$  as the unique series solution in  $\mathbf{s}$  to the following equation

$$w = X + \sum_{k \geq 1} k \bar{s}_k w^k, \quad w(X, \mathbf{s}) = X + \dots \quad (72)$$

The following theorem was recently obtained in [17].

**Theorem [17].** *The genus zero part of GUE free energy with even couplings has the expression*

$$\begin{aligned} \mathcal{F}_0 &= \frac{w^2}{4} - X w + \sum_{k \geq 1} \bar{s}_k \left( X w^k - \frac{k}{k+1} w^{k+1} \right) \\ &\quad + \frac{1}{2} \sum_{k_1, k_2 \geq 1} \frac{k_1 k_2}{k_1 + k_2} \bar{s}_{k_1} \bar{s}_{k_2} w^{k_1+k_2} + \frac{X^2}{2} \log w. \end{aligned} \quad (73)$$

For  $g \geq 1$ , there exist functions  $F_g(z_0, z_1, \dots, z_{3g-2})$ ,  $g \geq 1$  of independent variables  $z_0, z_1, z_2, \dots, z_{3g-2}$  such that

$$\mathcal{F}_g(X, \mathbf{s}) = F_g \left( u(X, \mathbf{s}), \frac{\partial u(X, \mathbf{s})}{\partial X}, \dots, \frac{\partial^{3g-2} u(X, \mathbf{s})}{\partial X^{3g-2}} \right), \quad g \geq 1. \quad (74)$$

Here

$$u(X, \mathbf{s}) := \frac{\partial^2 \mathcal{F}_0(X, \mathbf{s})}{\partial X^2} = \log w(X, \mathbf{s}).$$

We now specialize the  $\bar{s}_k$  to

$$\bar{s}_k = \frac{s}{k \cdot k!} \quad (k \geq 1), \quad (75)$$

where  $s$  is a parameter. (At the end of the argument we will further specialize to  $X = s - 1$ . One can also consider the more general specialization  $X = s + a$ , where  $a$  is an arbitrary nonzero parameter, in which case  $\mathcal{F}_g$  has *two* singularities in the complex  $z$ -plane. This case will be treated in a subsequent publication.) Substituting (75) into equation (72) gives the equation

$$w = X + s(e^w - 1)$$

for the power series  $w = w(X, s)$ . We can rewrite this as  $w = X - e^{s/e} + T(se^{X-s})$ , where  $T(z)$  is the Lambert series (28). It follows immediately that the function  $u(X, s) = \log w(X, s)$  and its  $X$ -derivatives specialize under  $s = ez, X = s - 1$  to

$$u(ez - 1, ez) = \log(T - 1), \quad \frac{\partial^m u}{\partial X^m}(ez - 1, ez) = \frac{Q_m(T)}{(1 - T)^{2m}} \quad (m \geq 1), \quad (76)$$

where  $z = Te^{-T}$  as usual and where  $Q_m(T)$  is a polynomial of degree  $m - 1$  in  $T$  with integer coefficients, with  $Q_1 = -1$ . Then we have

$$\begin{aligned} \mathcal{F} &= \mathcal{F}(\epsilon, z) := \mathcal{F}(X, \mathbf{s}; \epsilon) \Big|_{\bar{s}_k = \frac{ez}{k \cdot k!}, X = ez - 1} \\ &= \frac{(ez - 1)^2}{2\epsilon^2} \left( \log(ez - 1) - \frac{3}{2} \right) - \frac{1}{12} \log(ez - 1) + \zeta'(-1) \\ &\quad + \sum_{g \geq 2} \epsilon^{2g-2} \frac{B_{2g}}{4g(g-1)(ez-1)^{2g-2}} \\ &\quad + \sum_{\Gamma} \epsilon^{2g(\Gamma)-2} \frac{(ez)^{|V(\Gamma)|}}{\#\text{Sym } \Gamma} (ez - 1)^{|F(\Gamma)|} \prod_{i=1}^{|V(\Gamma)|} \frac{1}{\frac{\text{val}_i(\Gamma)}{2} \cdot (\frac{\text{val}_i(\Gamma)}{2})!} \\ &=: \sum_{g \geq 0} \epsilon^{2g-2} \mathcal{F}_g(z). \end{aligned} \quad (77)$$

Here the last summation in (77) is taken over all connected oriented ribbon graphs  $\Gamma$  with labelled half-edges and unlabelled vertices of even valencies,  $V(\Gamma)$  denotes the set of vertices of  $\Gamma$ ,  $\text{val}_1(\Gamma), \dots, \text{val}_{|V(\Gamma)|}(\Gamma)$  the valencies,  $E(\Gamma)$  the set of edges,  $F(\Gamma)$  the set of faces, and  $g(\Gamma)$  the genus.

**Theorem 6.** For  $g = 0, 1, 2, 3$ ,  $\mathcal{F}_g(z)$  have the following explicit expressions

$$\begin{aligned} \mathcal{F}_0(z) &= \frac{T^2}{4} - \frac{T}{2} - \frac{3}{4} + \frac{1}{2} \log(T - 1) - (\Gamma(0, 2 - 2T) + \gamma + i\pi + \log 2) \frac{T^2}{2} e^{2-2T} \\ &\quad + (\Gamma(0, 1 - T) + 1 + \gamma + i\pi) T e^{1-T}, \\ \mathcal{F}_1(z) &= \frac{1}{12} \log(-1) + \zeta'(-1) + \frac{1}{12} \log \frac{1}{(1 - T)^2}, \\ \mathcal{F}_2(z) &= -\frac{8T^3 + 43T^2 + 26T + 12}{2880(1 - T)^4}, \\ \mathcal{F}_3(z) &= \frac{32T^6 + 7392T^5 + 19953T^4 + 3668T^3 - 538T^2 + 1868T + 720}{725760(1 - T)^8}, \end{aligned}$$

where  $\gamma$  denotes Euler's constant,  $\Gamma(0, x)$  denotes the incomplete Gamma function, and  $z = Te^{-T}$ . Moreover,  $\forall g \geq 2$ ,  $\mathcal{F}_g(z)$  belongs to the Lambert space, more

precisely,

$$\mathcal{F}_g(z) = \sum_{i=g-1}^{4g-4} \frac{\mu_{g,i}}{(1-T)^i}, \tag{78}$$

where  $\mu_{g,i}$  are rational numbers satisfying  $\sum_{i=g-1}^{4g-4} \mu_{g,i} = \frac{B_{2g}}{2g(2g-2)}$ .

*Proof.* The validity of (78) is ensured by equations (74) and (76) together with the vanishing of  $\frac{\partial F_g}{\partial z_0}$  for  $g \geq 2$  and homogeneity statements for  $F_g$  that can be proved by using the Hodge–GUE correspondence [17], [14]. To obtain the explicit expressions of  $\mathcal{F}_2$  and  $\mathcal{F}_3$ , one can use the formulas of  $F_2, F_3$  in [17] with the specialization (76). So far, we do not know formulas for the top and the bottom coefficients  $\mu_{g,4g-4}$  and  $\mu_{g,g-1}$  for general values of  $g$ .  $\square$

**3.3. Hodge integrals.** In this final section we consider the evaluation of certain integrals over moduli spaces of curves that turn out to give a huge class of further sequences belonging to the Lambert space, and generalizing the original Hurwitz numbers studied in Section 1.

Let  $\overline{\mathcal{M}}_{g,n}$  denote the Deligne–Mumford moduli space of stable algebraic curves of genus  $g$  with  $n$  distinct marked points. Denote by  $\mathcal{L}_i$  the  $i$ -th tautological line bundle on  $\overline{\mathcal{M}}_{g,n}$ , by  $\mathbb{E}$  the rank  $g$  Hodge bundle, and by  $\psi_i$  the first Chern class  $c_1(\mathcal{L}_i)$ ,  $i = 1, \dots, n$ . We define the *Hodge free energy* as the generating series

$$\mathcal{H}^{\text{Hodge}} = \mathcal{H}^{\text{Hodge}}(\mathbf{t}; \mathbf{x}; \epsilon) = \sum_{g \geq 0} \mathcal{H}_g^{\text{Hodge}}(\mathbf{t}; \mathbf{x}) \epsilon^{2g-2}$$

of genus  $g$  parts defined by

$$\mathcal{H}_g^{\text{Hodge}}(\mathbf{t}; \mathbf{x}) = \sum_{n \geq 0} \sum_{i_1, \dots, i_n \geq 0} \frac{t_{i_1} \cdots t_{i_n}}{n!} \int_{\overline{\mathcal{M}}_{g,n}} \exp\left(\sum_{j \geq 0} x_{2j-1} \text{ch}_{2j-1}(\mathbb{E})\right) \psi_1^{i_1} \cdots \psi_n^{i_n}.$$

Here  $t_0, t_1, t_2, \dots$  are indeterminates (“coupling constants”),  $x_1, x_3, \dots$  are parameters,  $\mathbf{t} = (t_0, t_1, \dots)$ , and  $\mathbf{x} = (x_1, x_3, \dots)$ . According to Mumford [39], the even components of the Chern character of  $\mathbb{E}$  vanish. Hence  $\mathcal{H}^{\text{Hodge}}$  gives the generating series of Hodge integrals of the most general type.

We now define  $v = v(t_0, t_1, t_2, \dots)$  as the unique power series solution to

$$\sum_{i \geq 0} \frac{t_i}{i!} v^i = v. \tag{79}$$

By the Lagrange inversion formula,  $v$  has the following explicit expansion

$$v = \sum_{k \geq 1} \frac{1}{k} \sum_{p_1 + \dots + p_k = k-1} \frac{t_{p_1}}{p_1!} \cdots \frac{t_{p_k}}{p_k!}.$$

We also set  $v_m = \partial_{t_0}^m v$  for all  $m \geq 0$ . The following theorem for  $\mathcal{H}_g^{\text{Hodge}}$  was proved in [13].

**Theorem** [13], [15]. *We have*

$$\begin{aligned} \mathcal{H}_0^{\text{Hodge}}(\mathbf{t}; \mathbf{x}) &= \frac{v^3}{6} - \sum_{i \geq 0} t_i \frac{v^{i+2}}{i!(i+2)} + \frac{1}{2} \sum_{i,j \geq 0} t_i t_j \frac{v^{i+j+1}}{(i+j+1)!j!}, \\ \mathcal{H}_1^{\text{Hodge}}(\mathbf{t}; \mathbf{x}) &= \frac{1}{24} \log v_1 + \frac{x_1}{24} v. \end{aligned}$$

For  $g \geq 2$ , there exist algorithmically computable functions  $H_g^{\text{Hodge}}(z_1, \dots, z_{3g-3}; x_1, x_3, \dots, x_{2g-1})$ , rational in  $z_1$  and polynomial in all the other variables, such that

$$H_g^{\text{Hodge}}(\mathbf{t}; \mathbf{x}) = H_g^{\text{Hodge}}(v_1, \dots, v_{3g-3}; x_1, x_3, \dots, x_{2g-1}). \tag{80}$$

Moreover, the functions  $H_g^{\text{Hodge}}$  satisfy the homogeneity conditions

$$\sum_{m=1}^{3g-3} m z_m \frac{\partial H_g^{\text{Hodge}}}{\partial z_m} = (2g-2) H_g^{\text{Hodge}}, \tag{81}$$

$$\sum_{m=2}^{3g-3} (m-1) z_m \frac{\partial H_g^{\text{Hodge}}}{\partial z_m} + \sum_{j=1}^g (2j-1) x_j \frac{\partial H_g^{\text{Hodge}}}{\partial x_j} = (3g-3) H_g^{\text{Hodge}}. \tag{82}$$

We will be interested in the specialization of  $\mathcal{H}_g^{\text{Hodge}}$  when all variables  $t_i$  have the same value  $z$ , in which case we write simply  $z$  for  $\mathbf{t}$ . For this value of  $\mathbf{t}$  we have

$$\mathcal{H}^{\text{Hodge}}(z; \mathbf{x}; \epsilon) = \sum_{g,n \geq 0} c_{g,n}(\mathbf{x}) \epsilon^{2g-2} z^n, \quad \mathcal{H}_g^{\text{Hodge}}(z; \mathbf{x}) = \sum_{n \geq 0} c_{g,n}(\mathbf{x}) z^n$$

with

$$c_{g,n}(\mathbf{x}) := \frac{1}{n!} \sum_{i_1, \dots, i_n \geq 0} \int_{\mathcal{M}_{g,n}} \exp\left(\sum_{j \geq 0} x_{2j-1} \text{ch}_{2j-1}(\mathbb{E})\right) \psi_1^{i_1} \cdots \psi_n^{i_n}.$$

To compute  $\mathcal{H}_g^{\text{Hodge}}(z; \mathbf{x})$ , we first make the smaller specialization when  $t_i = t_1$  for  $i \geq 1$  but  $t_0$  is still an independent variable, and will then specialize to  $t_0 = t_1 = z$  at the end. (Compare Section 2, where we kept only two variables  $s$  and  $t$  of the infinitely many variables  $t_p^\alpha$  and specialized to  $s = 0$  at the end.) Then equation (79) simplifies to

$$t_0 + t_1 (e^v - 1) = v.$$

This means that if we set  $t_1 e^{t_0-t_1} = T e^{-T}$ , then  $v = v(t_0, t_1, t_1, \dots)$  is given by  $v = t_0 - t_1 + T$ . It follows easily that  $v_m(z) = v_m(z, z, \dots)$  is given by

$$v_m(z) = D^m T(z) + \delta_{m,1} = \overline{\alpha_{m-1}} + \delta_{m,1} \quad (m \geq 0), \tag{83}$$

where  $D = z \frac{d}{dz}$  as usual and  $\overline{\alpha_{m-1}}$  is defined in Section 1.4. This means a very simple explicit formula for  $v_m(z)$ : the coefficient of  $z^d$  in  $v_m(z)$  is  $d^{d+m-1}$  for all  $d \geq 1$ .

**Theorem 7.** *The power series  $\mathcal{H}_g^{\text{Hodge}}(z; \mathbf{x}) + \delta_{g,0}(z + z^2/4)$  belongs to the extended Lambert space for every  $g \geq 0$  and every  $\mathbf{x} = (x_1, \dots, x_{2g-1})$ . These power series*

are given explicitly for  $g = 0, 1, 2$  by

$$\mathcal{H}_0^{\text{Hodge}}(z; \mathbf{x}) + \frac{z^2}{4} + z = \frac{T^3}{6} - \frac{3T^2}{4} + T, \tag{84}$$

$$\mathcal{H}_1^{\text{Hodge}}(z; \mathbf{x}) = \frac{1}{24} \log \frac{1}{1-T} + \frac{T}{24} x_1, \tag{85}$$

$$\begin{aligned} \mathcal{H}_2^{\text{Hodge}}(z; \mathbf{x}) = & \frac{2x_1^3 - 42x_1^2 + 78x_1 - x_3 - 24}{34560(T-1)^2} + \frac{-7x_1^2 + 38x_1 - 31}{5760(T-1)^3} \\ & + \frac{5x_1 - 11}{1152(T-1)^4} - \frac{7}{1440(T-1)^5}. \end{aligned} \tag{86}$$

In general, for  $g \geq 2$ ,  $\mathcal{H}_g^{\text{Hodge}}(z; \mathbf{x})$  has the form

$$\mathcal{H}_g^{\text{Hodge}}(z; \mathbf{x}) = \sum_{i=2g-2}^{5g-5} \frac{\ell_{g,i}(\mathbf{x})}{(1-T)^i}, \quad g \geq 2, \tag{87}$$

where  $\ell_g(\mathbf{x})$  are polynomials in  $x_1, x_3, \dots, x_{2g-1}$ . Moreover,  $\ell_{g,5g-5}(\mathbf{x})$  is independent of  $\mathbf{x}$  and is equal to  $24^{-g}c_g/((5g-3)(5g-5))$ , where  $c_g$  is defined by (7) or (8).

*Proof.* The formulas (84)–(86) can be obtained from the algorithm developed in [13] with the particular  $v_m$  given by (83). The formula (87) follows from (80), (83) and the homogeneity conditions (81), (82). The fact that  $\ell_{g,5g-5}$  does not depend on  $\mathbf{x}$  is due to the dimension-degree matching.  $\square$

We note that the first statement of Theorem 7 is a particular case of the Theorem 1 in [46], but no algorithm is given there to compute the Hodge integrals.

Noting that

$$1 - \lambda_1 + \lambda_2 - \dots + (-1)^g \lambda_g = \exp\left(-\sum_{j \geq 1} (2j-2)! \text{ch}_{2j-1}\right)$$

we have from the ELSV formula (39) that

$$\mathcal{H}_g(z) := \sum_{d \geq 0} h_{g,d} z^d = \mathcal{H}_g^{\text{Hodge}}(z; -0!, -2!, -4!, \dots, -(2g-2)!).$$

Thus the generating series  $\mathcal{H}_g(z; \mathbf{x})$  generalizes the power series  $\mathcal{H}_g(z)$  of Section 1, and Theorem 7 generalizes Theorem 3. (For instance, for  $g = 2$  we verify easily that under  $x_1 = -1, x_3 = -2$ , the formula (86) becomes (33).) Combining the two theorems, we obtain

**Corollary.** For any fixed  $g \geq 0$ , and any fixed value  $\mathbf{x} = (x_1, \dots, x_{2g-1})$ , the function  $\mathcal{H}_g^{\text{Hodge}}(z; \mathbf{x})$  is analytic around  $z = 0$ , and has the dominant singularity at  $z = e^{-1}$ . As  $n \rightarrow \infty$

$$c_{g,n}(\mathbf{x}) \sim h_{g,n}.$$

## REFERENCES

- [1] V. I. Arnold, *Topological classification of complex trigonometric polynomials and the combinatorics of graphs with an identical number of vertices and edges*, Funktsional. Anal. i Prilozhen. **30** (1996), no. 1, 1–17, 96 (Russian). MR [1387484](#). English translation: Funct. Anal. Appl. **30** (1996), no. 1, 1–14.
- [2] M. Bertola, B. Dubrovin, and D. Yang, *Correlation functions of the KdV hierarchy and applications to intersection numbers over  $\overline{\mathcal{M}}_{g,n}$* , Phys. D **327** (2016), 30–57. MR [3505204](#)
- [3] D. Bessis, C. Itzykson, and J. B. Zuber, *Quantum field theory techniques in graphical enumeration*, Adv. in Appl. Math. **1** (1980), no. 2, 109–157. MR [603127](#)
- [4] V. Bouchard and M. Mariño, *Hurwitz numbers, matrix models and enumerative geometry*, Proc. Sympos. Pure Math., vol. 78, Amer. Math. Soc., Providence, RI, 2008, pp. 263–283. MR [2483754](#)
- [5] A. Buryak, F. Janda, and R. Pandharipande, *The hypergeometric functions of the Faber–Zagier and Pixton relations*, preprint [arXiv:1502.05150](#) [math.AG].
- [6] N. Caporaso, L. Griguolo, M. Mariño, S. Pasquetti, and D. Seminara, *Phase transitions, double-scaling limit, and topological strings*, Phys. Rev. D **75** (2007), no. 4, 046004, 24. MR [2304432](#)
- [7] G. Carlet, B. Dubrovin, and Y. Zhang, *The extended Toda hierarchy*, Mosc. Math. J. **4** (2004), no. 2, 313–332, 534. MR [2108440](#)
- [8] D. Chen, M. Moeller, and Z. Don, *Quasimodularity and large genus limits of Siegel–Veech constants*, preprint [arXiv:1606.04065](#) [math.NT].
- [9] B. Dubrovin, *Geometry of 2D topological field theories*, Integrable systems and quantum groups (Montecatini Terme, 1993), Lecture Notes in Math., vol. 1620, Springer, Berlin, 1996, pp. 120–348. MR [1397274](#)
- [10] B. Dubrovin, *Hamiltonian perturbations of hyperbolic PDEs: from classification results to the properties of solutions*, New trends in mathematical physics. Selected contributions of the XVth international congress on mathematical physics (V. Sidoravicius, Ed.), Springer, 2009, pp. 231–276.
- [11] B. Dubrovin, *Gromov–Witten invariants and integrable hierarchies of topological type*, Topology, geometry, integrable systems, and mathematical physics, Amer. Math. Soc. Transl. Ser. 2, vol. 234, Amer. Math. Soc., Providence, RI, 2014, pp. 141–171. MR [3307147](#)
- [12] B. Dubrovin, *Symplectic field theory of a disk, quantum integrable systems, and Schur polynomials*, Ann. Henri Poincaré **17** (2016), no. 7, 1595–1613. MR [3510464](#)
- [13] B. Dubrovin, S.-Q. Liu, D. Yang, and Y. Zhang, *Hodge integrals and tau-symmetric integrable hierarchies of Hamiltonian evolutionary PDEs*, Adv. Math. **293** (2016), 382–435. MR [3474326](#)
- [14] B. Dubrovin, S.-Q. Liu, D. Yang, and Y. Zhang, *Hodge-GUE correspondence and the discrete KdV equation*, preprint [arXiv:1612.02333](#) [math-ph].
- [15] B. Dubrovin and D. Yang, *Remarks on intersection numbers and integrable hierarchies. I. Quasi-triviality*, to appear.
- [16] B. Dubrovin and D. Yang, *Generating series for GUE correlators*, Lett. Math. Phys. **107** (2017), 1971–2012. Preprint version: [arXiv:1604.07628](#) [math-ph].
- [17] B. Dubrovin and D. Yang, *On cubic Hodge integrals and random matrices*, Commun. Number Theory and Phys. **11** (2017), no. 2, 311–336. preprint [arXiv:1606.03720](#) [math-ph].
- [18] B. Dubrovin and Y. Zhang, *Normal forms of hierarchies of integrable PDEs, Frobenius manifolds and Gromov–Witten invariants*, preprint [arXiv:math/0108160](#) [math.DG].
- [19] B. Dubrovin and Y. Zhang, *Virasoro symmetries of the extended Toda hierarchy*, Comm. Math. Phys. **250** (2004), no. 1, 161–193. MR [2092034](#)
- [20] P. Dunin-Barkowski, M. Kazarian, N. Orantin, S. Shadrin, and L. Spitz, *Polynomiality of Hurwitz numbers, Bouchard–Mariño conjecture, and a new proof of the ELSV formula*, Adv. Math. **279** (2015), 67–103. MR [3345179](#)
- [21] T. Ekedahl, S. Lando, M. Shapiro, and A. Vainshtein, *Hurwitz numbers and intersections on moduli spaces of curves*, Invent. Math. **146** (2001), no. 2, 297–327. MR [1864018](#)

- [22] C. Faber, *A conjectural description of the tautological ring of the moduli space of curves*, Moduli of curves and abelian varieties, Aspects Math., E33, Friedr. Vieweg, Braunschweig, 1999, pp. 109–129. MR [1722541](#)
- [23] C. Faber and R. Pandharipande, *Hodge integrals and Gromov–Witten theory*, Invent. Math. **139** (2000), no. 1, 173–199. MR [1728879](#)
- [24] P. Flajolet and R. Sedgewick, *Analytic combinatorics*, Cambridge University Press, Cambridge, 2009. MR [2483235](#)
- [25] S. Garoufalidis, A. Its, A. Kapaev, and M. Mariño, *Asymptotics of the instantons of Painlevé I*, Int. Math. Res. Not. IMRN (2012), no. 3, 561–606. MR [2885983](#)
- [26] E. Getzler, A. Okounkov, and R. Pandharipande, *Multipoint series of Gromov–Witten invariants of  $\mathbb{C}P^1$* , Lett. Math. Phys. **62** (2002), no. 2, 159–170. MR [1952124](#)
- [27] I. P. Goulden and D. M. Jackson, *The number of ramified coverings of the sphere by the double torus, and a general form for higher genera*, J. Combin. Theory Ser. A **88** (1999), no. 2, 259–275. MR [1723797](#)
- [28] I. P. Goulden, D. M. Jackson, and R. Vakil, *The Gromov–Witten potential of a point, Hurwitz numbers, and Hodge integrals*, Proc. London Math. Soc. (3) **83** (2001), no. 3, 563–581. MR [1851082](#)
- [29] J. Harer and D. Zagier, *The Euler characteristic of the moduli space of curves*, Invent. Math. **85** (1986), no. 3, 457–485. MR [848681](#)
- [30] A. Hurwitz, *Ueber Riemann’sche Flächen mit gegebenen Verzweigungspunkten*, Math. Ann. **39** (1891), no. 1, 1–60. MR [1510692](#)
- [31] A. Hurwitz, *Ueber die Anzahl der Riemann’schen Flächen mit gegebenen Verzweigungspunkten*, Math. Ann. **55** (1901), no. 1, 53–66. MR [1511135](#)
- [32] C. Itzykson and J.-B. Zuber, *Combinatorics of the modular group. II. The Kontsevich integrals*, Internat. J. Modern Phys. A **7** (1992), no. 23, 5661–5705. MR [1180858](#)
- [33] N. Joshi and A. V. Kitaev, *On Boutroux’s tritronquée solutions of the first Painlevé equation*, Stud. Appl. Math. **107** (2001), no. 3, 253–291. MR [1854431](#)
- [34] A. A. Kapaev, *Quasi-linear stokes phenomenon for the Painlevé first equation*, J. Phys. A **37** (2004), no. 46, 11149–11167. MR [2101932](#)
- [35] M. Kazarian, *KP hierarchy for Hodge integrals*, Adv. Math. **221** (2009), no. 1, 1–21. MR [2509319](#)
- [36] S. K. Lando and A. K. Zvonkin, *Graphs on surfaces and their applications*, Encyclopaedia of Mathematical Sciences, vol. 141, Springer-Verlag, Berlin, 2004. MR [2036721](#)
- [37] Yu. I. Manin, *Generating functions in algebraic geometry and sums over trees*, The moduli space of curves (Texel Island, 1994), Progr. Math., vol. 129, Birkhäuser Boston, Boston, MA, 1995, pp. 401–417. MR [1363064](#)
- [38] A. Morozov and S. Shakirov, *Exact 2-point function in Hermitian matrix model*, J. High Energy Phys. (2009), no. 12, 003, 33. MR [2593032](#)
- [39] D. Mumford, *Towards an enumerative geometry of the moduli space of curves*, Arithmetic and geometry, Vol. II, Progr. Math., vol. 36, Birkhäuser Boston, Boston, MA, 1983, pp. 271–328. MR [717614](#)
- [40] A. Okounkov, *Toda equations for Hurwitz numbers*, Math. Res. Lett. **7** (2000), no. 4, 447–453. MR [1783622](#)
- [41] A. Okounkov and R. Pandharipande, *Gromov–Witten theory, Hurwitz theory, and completed cycles*, Ann. of Math. (2) **163** (2006), no. 2, 517–560. MR [2199225](#)
- [42] A. Okounkov and R. Pandharipande, *Gromov–Witten theory, Hurwitz numbers, and matrix models*, Algebraic geometry—Seattle 2005. Part 1, Proc. Sympos. Pure Math., vol. 80, Amer. Math. Soc., Providence, RI, 2009, pp. 325–414. MR [2483941](#)
- [43] R. Pandharipande, *The Toda equations and the Gromov–Witten theory of the Riemann sphere*, Lett. Math. Phys. **53** (2000), no. 1, 59–74. MR [1799843](#)
- [44] J. Zhou, *On topological 1D gravity. I*, preprint [arXiv:1412.1604](#) [math-ph].
- [45] J. Zhou, *On a mean field theory of topological 2D gravity*, preprint [arXiv:1503.08546](#) [math.AG].

- [46] D. Zvonkine, *An algebra of power series arising in the intersection theory of moduli spaces of curves and in the enumeration of ramified coverings of the sphere*, preprint [arXiv:math/0403092](https://arxiv.org/abs/math/0403092) [math.AG].

B.D.: SISSA, VIA BONOMEA 265, TRIESTE 34136, ITALY  
*E-mail address:* [dubrovin@sissa.it](mailto:dubrovin@sissa.it)

D.Y.: MAX-PLANCK-INSTITUT FÜR MATHEMATIK, VIVATSGASSE 7, BONN 53111, GERMANY  
*E-mail address:* [diyang@mpim-bonn.mpg.de](mailto:diyang@mpim-bonn.mpg.de)

D.Z.: MAX-PLANCK-INSTITUT FÜR MATHEMATIK, VIVATSGASSE 7, BONN 53111, GERMANY  
*E-mail address:* [dbz@mpim-bonn.mpg.de](mailto:dbz@mpim-bonn.mpg.de)