Matrix Group Integrals, Surfaces, and Mapping Class Groups I: $\mathcal{U}(n)$

Michael Magee* and Doron Puder†

May 3, 2019

Abstract

Since the 1970's, physicists and mathematicians who study random matrices in the GUE or GOE models are aware of intriguing connections between integrals of such random matrices and enumeration of graphs on surfaces. We establish a new aspect of this theory: for random matrices sampled from the group $\mathcal{U}(n)$ of unitary matrices.

More concretely, we study measures induced by free words on $\mathcal{U}(n)$. Let $\mathbf{F}_r$ be the free group on r generators. To sample a random element from $\mathcal{U}(n)$ according to the measure induced by $w \in \mathbf{F}_r$, one substitutes the r letters in w by r independent, Haar-random elements from $\mathcal{U}(n)$. The main theme of this paper is that every moment of this measure is determined by families of pairs (Σ, f), where Σ is an orientable surface with boundary, and f is a map from Σ to the bouquet of r circles, which sends the boundary components of Σ to powers of w. A crucial role is then played by Euler characteristics of subgroups of the mapping class group of Σ.

As corollaries, we obtain asymptotic bounds on the moments, we show that the measure on $\mathcal{U}(n)$ bears information about the number of solutions to the equation $[u_1, v_1] \cdots [u_g, v_g] = w$ in the free group, and deduce that one can "hear" the stable commutator length of a word through its unitary word measures.

Contents

*M. Magee was partially supported by the N.S.F. via grants DMS-1128155 and DMS-1701357.

†D. Puder was partially supported by the Rothschild Fellowship, N.S.F. via grant DMS-1128155, and I.S.F. via grant 1071/16.

1 Introduction

Let $\mathcal{U}(n)$ denote the group of $n \times n$ unitary complex matrices, and let $\mathbf{F}_r$ denote the free group on r generators with fixed basis (free generating set) $B = \{x_1, \ldots, x_r\}$. For a word $w \in \mathbf{F}_r$, we define the w-measure on $\mathcal{U}(n)$ as the push-forward of the Haar measure on $\mathcal{U}(n)^r$ through the word map $w\colon \mathcal{U}(n)^r \to \mathcal{U}(n)$. In plain terms, assume that $w = x_{i_1}^{\varepsilon_1} \cdots x_{i_m}^{\varepsilon_m}$. To sample a random element from $\mathcal{U}(n)$ by the w-measure, sample r independent Haar-random elements $A_1, \ldots, A_r \in \mathcal{U}(n)$ and evaluate $w(A_1, \ldots, A_r) = A_{i_1}^{\varepsilon_1} \cdots A_{i_m}^{\varepsilon_m} \in \mathcal{U}(n)$. $\mathcal{U}(n)$ $\mathbf{F}_r$ B

The motivation to study w-measures on unitary groups or on compact groups in general originates in questions revolving around random walks on these groups, in the study of representation varieties, in problems in the theory of Free Probability, and in challenges in the study of free groups. However, as the current paper shows, the study of w-measures is interesting for its own sake and reveals deep and surprising connections with other mathematical concepts. See also [PP15].

Expected trace

We study word measures by considering their moments, and more particularly the expected product of traces. For every $\ell \in \mathbb{N}_{\geq 1}$ and $w_1, \ldots, w_\ell \in \mathbf{F}_r$, consider the quantity

$$\mathcal{T}r_{w_1,\ldots,w_\ell}(n) \stackrel{\text{def}}{=} \int_{A_1,\ldots,A_r \in \mathcal{U}(n)} \mathrm{tr}\left(w_1(A_1,\ldots,A_r)\right) \cdot \ldots \cdot \mathrm{tr}\left(w_\ell(A_1,\ldots,A_r)\right) d\mu \tag{1.1}$$

where $A_1, \ldots, A_r \in \mathcal{U}(n)$ are independent Haar-random unitary matrices[1]. The development of "Weingarten calculus" for computing integrals on $\mathcal{U}(n)$ [Wei78, Xu97, Col03, CŚ06] leads readily to the following result:

Proposition 1.1. *Let $\ell \in \mathbb{N}_{\geq 1}$ and $w_1, \ldots, w_\ell \in \mathbf{F}_r$. Then for large enough n, the quantity $\mathcal{T}r_{w_1,\ldots,w_\ell}(n)$ is given by a rational expression in n with rational coefficients, namely, by an element of $\mathbb{Q}(n)$.*

Here "large enough n" means that $n \geq \max_{x \in B} L_x$, where L_x is the total number of instances of x^{+1} in the words $w_1, \ldots, w_\ell$.

[1] Let us mention that the w-measure on $\mathcal{U}(n)$ is completely determined by moments of this type where the words are taken to be powers of w: $\mathcal{T}r_{w^{\alpha_1},\ldots,w^{\alpha_\ell}}(n)$ with $\alpha_1, \ldots, \alpha_\ell \in \mathbb{Z}$. See, for example, [MP15, Section 2.2]. (We comment about the pre-print [MP15] in Remark 1.15.)

ℓ	$w_1, \ldots, w_\ell$	$\mathcal{T}r_{w_1,\ldots,w_\ell}(n)$	Laurent Series
	$[x,y]$	$\frac{1}{n}$	$\frac{1}{n}$
	$[x^3,y]$	$\frac{3}{n}$	$\frac{3}{n}$
1	$[x,y]^2$	$\frac{-4}{n^3-n}$	$\frac{-4}{n^3} + \frac{-4}{n^5} + \frac{-4}{n^7} + \cdots$
	$[x,y]^3$	$\frac{9(n^2+4)}{n^5-5n^3+4n}$	$\frac{9}{n^3} + \frac{81}{n^5} + \frac{369}{n^7} + \cdots$
	$[x,y][x,z]$	0	0
	$[x,y][x,z][x,t]$	0	0
	$x^2y^2, xy^{-3}x^{-3}y$	$\frac{4(n^2-5)}{n^4-5n^2+4}$	$\frac{4}{n^2} + \frac{0}{n^4} + \frac{-16}{n^6} + \frac{-80}{n^8} + \cdots$
2	w, w^{-1} for $w = x^2yxy^{-1}$	1	1
	w, w^{-1} for $w = x^2y^2xy^{-1}$	$\frac{n^4-5n^2}{n^4-5n^2+4}$	$1 + \frac{0}{n^2} + \frac{-4}{n^4} + \frac{-20}{n^6} + \cdots$

Table 1: Some examples for the rational expression for $\mathcal{T}r_{w_1,\ldots,w_\ell}(n)$ and (the beginning of) its Laurent series expansion. All these examples contain words in $\mathbf{F}_4$ with generators $\{x,y,z,t\}$. The notation $[x,y]$ is for the commutator $xyx^{-1}y^{-1}$.

In Section 2 we give explicit combinatorial formulas for $\mathcal{T}r_{w_1,\ldots,w_\ell}(n)$, and the main innovation here is the emergence of surfaces in these formulas. In Table 1 we list some examples[2] for these rational expressions for concrete words.

The main theme of the current paper is the interpretation of these expressions for $\mathcal{T}r_{w_1,\ldots,w_\ell}(n)$ in terms of properties of w. We explain their degree and their leading coefficient. More generally, we show how the entire Laurent series for $\mathcal{T}r_{w_1,\ldots,w_\ell}(n)$ is determined by natural objects related to $w_1, \ldots, w_\ell$.

Extending maps from circles to surfaces

Our main result, Theorem 1.7 below, states that the expressions for $\mathcal{T}r_{w_1,\ldots,w_\ell}(n)$ can be described in terms of certain surfaces and maps. Roughly, consider a bouquet of r circles $\bigvee^r S^1$ with fundamental group identified with $\mathbf{F}_r$. Now consider ℓ disjoint circles (one-spheres) $C_1, \ldots, C_\ell$ and a map $\bigvee^r S^1$

$$f_{w_1,\ldots,w_\ell} \colon C_1 \sqcup \ldots \sqcup C_\ell \to \bigvee^r S^1$$

sending C_i to a loop at the bouquet representing w_i. We now construct pairs (Σ, f) of an orientable surface Σ with ℓ boundary components together with a map $f \colon \Sigma \to \bigvee^r S^1$, so that the restriction of f to the boundary $\partial\Sigma$ is equal to $f_{w_1,\ldots,w_\ell}$. From this set one can fully recover the expressions for $\mathcal{T}r_{w_1,\ldots,w_\ell}(n)$. See Figure 1.1.

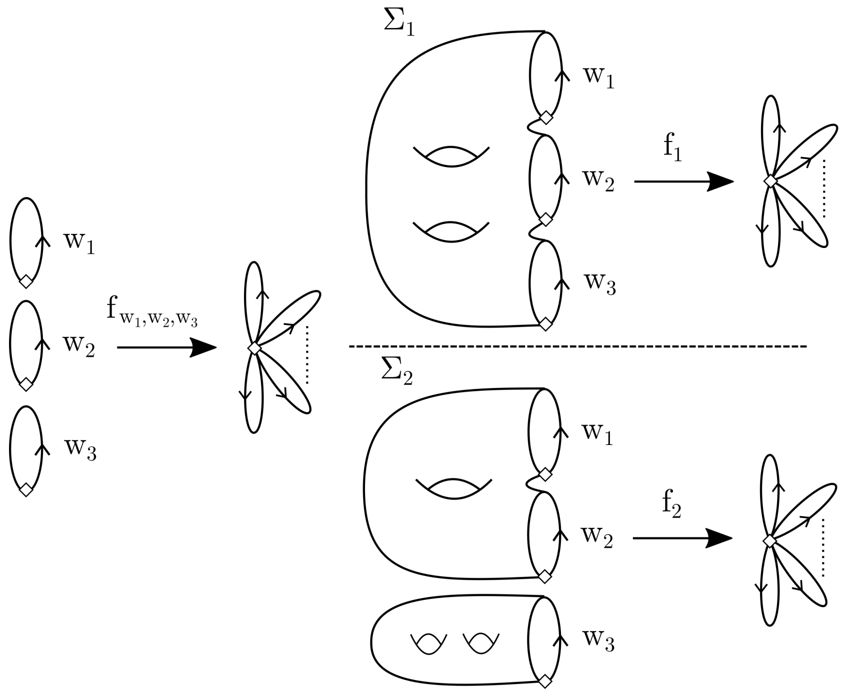

Figure 1.1: Two surfaces extending the map from three circles to $\bigvee^r S^1$ corresponding to the triple of words w_1, w_2, w_3.

More formally, identify the free group $\mathbf{F}_r$ with the fundamental group of $\bigvee^r S^1$, by orienting every circle in the bouquet and determining a bijection between the circles and the generators $x_1, \ldots, x_r$ of $\mathbf{F}_r$. Mark the wedge point by o. We have o

$$\mathbf{F}_r \cong \pi_1\left(\bigvee^r S^1, o\right).$$

Let $C_1 \sqcup \ldots \sqcup C_\ell$ be a disjoint union of ℓ oriented 1-spheres with a marked point $v_i \in C_i$ for every v $i = 1, \ldots, \ell$. The map $f_{w_1,\ldots,w_\ell} \colon C_1 \sqcup \ldots \sqcup C_\ell \to \bigvee^r S^1$ sends $v_1, \ldots, v_\ell$ to o, and the induced map on fundamental groups sends the loop at v_i around the oriented C_i to $[w_i]$.

Definition 1.2. Let Σ be a surface with ℓ boundary components $\partial\Sigma_1, \ldots, \partial\Sigma_\ell$ and a marked point $v_i \in \partial\Sigma_i$ in each boundary component. Let $f \colon \Sigma \to \bigvee^r S^1$ be a map to the bouquet. We say that v_i (Σ, f) is **admissible** for $w_1, \ldots, w_\ell \in \mathbf{F}_r$ if the following two conditions hold:

1. Σ is oriented and compact, with no closed connected components.
2. The restriction of f to the boundary of Σ is homotopic to $f_{w_1,\ldots,w_\ell}$ relative to the marked points $v_1, \ldots, v_\ell$. Namely, for every $i = 1, \ldots, \ell$,

$$f_*\left(\left[\overrightarrow{\partial_i\Sigma}\right]\right) = w_i \in \pi_1\left(\bigvee^r S^1, o\right),$$

where $\overrightarrow{\partial_i\Sigma}$ is the closed loop at v_i around $\partial_i\Sigma$ with orientation induced from the orientation of Σ.

In particular, we assume in the above definition that $f(v_i) = o$ for every $i = 1, \ldots, \ell$.

There is a natural equivalence relation between different admissible pairs: first, if $f_1, f_2 \colon \Sigma \to \bigvee^r S^1$ are homotopic relative to the marked points $v_1, \ldots, v_\ell$, then we think of (Σ, f_1) and (Σ, f_2) as equivalent. We denote by $[f]$ the homotopy class of f relative to $v_1, \ldots, v_\ell$. Second, there is a $[f]$

natural action of $\mathrm{MCG}(\Sigma)$, the mapping class group of Σ, on homotopy classes of maps $\Sigma \to \bigvee^r S^1$, and we define different maps in the same orbit to be equivalent (see Definition 1.3). Here, $\mathrm{MCG}(\Sigma)$ is defined as the group of homeomorphisms of Σ which fix the boundary $\partial\Sigma$ pointwise, modulo such homeomorphisms which are isotopic to the identity. The action of $\mathrm{MCG}(\Sigma)$ on homotopy classes of maps $\mathrm{MCG}(\Sigma)$

$$\{[f] \mid (\Sigma, f) \text{ admissible for } w_1, \ldots, w_\ell\}$$

is by precomposition: the action of $[\rho] \in \mathrm{MCG}(\Sigma)$ on $[f]$ results in $\left[f \circ \rho^{-1}\right]$. We gather these considerations in the following definition:

Definition 1.3. Let (Σ, f) and (Σ', f') be admissible for $w_1, \ldots, w_\ell$. They are **equivalent**, denoted $(\Sigma, f) \sim (\Sigma', f')$, if there is an orientation preserving homeomorphism $\rho \colon \Sigma \to \Sigma'$, such that for every $i = 1, \ldots, \ell$, $\rho(v_i) = v_i'$ and $f \simeq f' \circ \rho$ are homotopic relative to the marked points $v_1, \ldots v_\ell$. We denote by $[(\Sigma, f)]$ the equivalence class of (Σ, f). We denote the set of equivalence classes by $\mathsf{Surfaces}(w_1, \ldots, w_\ell)$: $(\Sigma,f)\sim(\Sigma',f')$ $[(\Sigma, f)]$ $\mathsf{Surfaces}(w_1,\ldots,w_\ell)$

$$\mathsf{Surfaces}(w_1, \ldots, w_\ell) \stackrel{\text{def}}{=} \{[(\Sigma, f)] \mid (\Sigma, f) \text{ is admissible for } w_1, \ldots, w_\ell\}.$$

The main goal of this paper is to show how one can read the terms of the Laurent series of $\mathcal{T}r_{w_1,\ldots,w_\ell}(n)$ from this set $\mathsf{Surfaces}(w_1, \ldots, w_\ell)$ of equivalence classes of pairs of surfaces and maps.

The L^2-Euler characteristic of stabilizers

The Laurent series of $\mathcal{T}r_{w_1,\ldots,w_\ell}(n)$ gets some contribution from every $[(\Sigma, f)] \in \mathsf{Surfaces}(w_1, \ldots, w_\ell)$. As stated in Theorem 1.7 below, this contribution is of the form $c \cdot n^\alpha$, where c and α are integers. The order of magnitude of the contribution is controlled by the Euler characteristic of the surface: $\alpha = \chi(\Sigma)$. However, to determine the integer coefficient c, an important role is played by the stabilizer of $[f]$ under the action of $\mathrm{MCG}(\Sigma)$, which we denote by $\mathrm{MCG}(f)$: $\mathrm{MCG}(f)$

$$\mathrm{MCG}(f) \stackrel{\text{def}}{=} \mathrm{MCG}(\Sigma)_{[f]}.$$

Note that by definition, the elements of $\mathrm{MCG}(\Sigma)$ permute homotopy classes of maps inside the same equivalence class $[(\Sigma, f)]$. Yet, occasionally, they may stabilize $[f]$, in the sense that $f \circ \rho$ and f are homotopic relative to $v_1, \ldots, v_\ell$. Given the class $[(\Sigma, f)]$, the stabilizer $\mathrm{MCG}(f)$ is defined up to conjugation.

The actual invariant of the stabilizer that appears in the contribution of $[(\Sigma, f)]$ to $\mathcal{T}r_{w_1,\ldots,w_\ell}(n)$ is its **L^2-Euler characteristic**. The L^2-Euler characteristic of a group is defined for groups with nice enough properties and can take any real value. It is the alternating sum of the von Neumann dimensions of the homology groups of a natural chain complex of modules over the group von Neumann algebra, as we explain in more detail in Section 4.1 below, and see [Lüc02]. Thus, to state our main result, we first need the following auxiliary theorem which is interesting for its own sake.

Theorem 1.4. *Let Σ be a compact orientable surface with no closed connected components. Let $f \colon \Sigma \to \bigvee^r S^1$ be a map. Then the stabilizer $\mathrm{MCG}(f) = \mathrm{MCG}(\Sigma)_{[f]}$ has a well-defined L^2-Euler characteristic. Moreover, this L^2-Euler Characteristic is an integer.*

Note that in the statement of the theorem it does not matter whether $[f]$ is the homotopy class of f relative to $\partial\Sigma$ or relative to some marked points in every boundary component - this nuance does not modify the action of $\mathrm{MCG}(\Sigma)$ on the homotopy classes of maps.

Theorem 1.4 can be strengthened in an important special case we now introduce:

Definition 1.5. A *null-curve* of (Σ, f) is a non-nullhomotopic simple closed curve γ in Σ with $f(\gamma)$ nullhomotopic in $\bigvee^r S^1$. A pair (Σ, f) is called *incompressible* if it admits no null-curves. It is called *compressible* otherwise. null-curve incompressible

If (Σ, f) is admissible for $w_1, \ldots, w_\ell$ and is compressible, then one can cut Σ along a null-curve, fill the two new boundary components with discs to obtain Σ' and extend f to a map $f' \colon \Sigma' \to \bigvee^r S^1$. If Σ' contains a closed component, remove it to obtain Σ'' and let f'' denote the restriction of f' to Σ''. The new pair (Σ'', f'') is admissible for $w_1, \ldots, w_\ell$ and satisfies $\chi(\Sigma'') \geq \chi(\Sigma') = \chi(\Sigma) + 2$, as the possibly closed component of Σ' cannot be a sphere. Thus, pairs (Σ, f) with Σ having maximal Euler characteristic are necessarily incompressible. When f is incompressible we have the following stronger version of Theorem 1.4:

Theorem 1.6. *Let Σ be a compact orientable surface with boundary in every connected component, and let $f \colon \Sigma \to \bigvee^r S^1$ be incompressible. Then the stabilizer*

$$\Gamma = \mathrm{MCG}(f) = \mathrm{MCG}(\Sigma)_{[f]}$$

admits a finite simplicial complex as a $K(\Gamma, 1)$-space. In particular, Γ has a well-defined Euler characteristic in the ordinary sense[3]*, which coincides with its L^2-Euler characteristic.*

Main result

Our main theorem shows that the Laurent expansion of $\mathcal{T}r_{w_1,\ldots,w_\ell}(n)$ is given by Euler characteristics of both the stabilizers of maps in $\mathsf{Surfaces}(w_1, \ldots, w_\ell)$ and also the Euler characteristics of the surfaces. When the L^2-Euler characteristic of a group Γ is defined, we denote it by $\chi^{(2)}(\Gamma)$. $\chi^{(2)}(\Gamma)$

Theorem 1.7 (Main Theorem)**.** *Let $w_1, \ldots, w_\ell \in \mathbf{F}_r$. For large enough*[4] *$n$,*

$$\mathcal{T}r_{w_1,\ldots,w_\ell}(n) = \sum_{[(\Sigma,f)] \in \mathsf{Surfaces}(w_1,\ldots,w_\ell)} \chi^{(2)}(\mathrm{MCG}(f)) \cdot n^{\chi(\Sigma)}. \tag{1.2}$$

Indeed, for any given exponent χ_0, there are only finitely many non-zero terms of order n^{χ_0}, namely, the set

$$\left\{ [(\Sigma, f)] \in \mathsf{Surfaces}(w_1, \ldots, w_\ell) \;\middle|\; \chi(\Sigma) = \chi_0 \text{ and } \chi^{(2)}(\mathrm{MCG}(f)) \neq 0 \right\}$$

is finite.

The last statement of the theorem explains why the theorem yields a well-defined coefficient for every term in the Laurent series of $\mathcal{T}r_{w_1,\ldots,w_\ell}(n)$. However, we do not know yet how to derive from this theorem the rationality of $\mathcal{T}r_{w_1,\ldots,w_\ell}(n)$, which we prove directly using Weingarten calculus - see Proposition 1.1 and Section 2. This rationality means that in a way we do not yet fully understand, the L^2-Euler characteristics of different pairs $[(\Sigma, f)] \in \mathsf{Surfaces}(w_1, \ldots, w_\ell)$ "know about each other" – see Question 4 in Section 6.

As an immediate corollary of Proposition 1.1 and Theorem 1.7, we get an asymptotic upper bound on $\mathcal{T}r_{w_1,\ldots,w_\ell}(n)$. Denote $\chi_{\max}(w_1,\ldots,w_\ell)$

[2] Every example in Table 1 satisfies that for every generator x_i, the total number of occurrences in $w_1, \ldots, w_\ell$ of x_i^{+1} is equal to the number of occurrences of x_i^{-1}. The reason is the simple fact that otherwise $\mathcal{T}r_{w_1,\ldots,w_\ell}(n)$ is constantly zero – see Claim 2.1 below.

[3] The "ordinary" Euler characteristic of a group is defined for a large class of groups of certain finiteness conditions – see [Bro82, Chapter IX]. The simplest case is when a group Γ admits a finite CW-complex as Eilenberg-MacLane space of type $K(\Gamma, 1)$, namely, a path-connected complex with fundamental group isomorphic to Γ and a contractible universal cover. In this case, the Euler characteristic of Γ coincides with the Euler characteristic of the $K(\Gamma, 1)$-space.

[4] As in Proposition 1.1, the equality (1.2) holds for every $n \geq \max_{x \in B} L_x$, where L_x is the total number of appearance of x^{+1} in the words $w_1, \ldots, w_\ell$. See also Section 2.

$$\chi_{\max}(w_1,\ldots,w_\ell) \stackrel{\text{def}}{=} \max\{\chi(\Sigma) \mid [(\Sigma,f)] \in \mathsf{Surfaces}(w_1,\ldots,w_\ell)\},$$

where $\chi_{\max}(w_1,\ldots,w_\ell) = -\infty$ if $\mathsf{Surfaces}(w_1,\ldots,w_\ell)$ is empty, which is equivalent to $w_1\cdots w_\ell \notin [\mathbf{F}_r,\mathbf{F}_r]$ - see Claims 2.1 and 2.12. A well-known fact going back at least to Culler [Cul81, Paragraph 1.1] is that $\mathrm{chi}(w) = 1 - 2\cdot \mathrm{cl}(w)$, where $\mathrm{cl}(w)$ is the commutator length of w, defined as[5] $\mathrm{cl}(w)$

$$\mathrm{cl}(w) \stackrel{\text{def}}{=} \min\{g \mid w = [u_1,v_1]\cdots[u_g,v_g] \text{ with } u_i,v_i \in \mathbf{F}_r\}.$$

Thus,

Corollary 1.8. *Let $w_1,\ldots,w_\ell \in \mathbf{F}_r$. Then*

$$\mathcal{T}r_{w_1,\ldots,w_\ell}(n) = O\left(n^{\chi_{\max}(w_1,\ldots,w_\ell)}\right). \tag{1.3}$$

In particular, for $w \in \mathbf{F}_r$,

$$\mathcal{T}r_w(n) = O\left(n^{\chi_{\max}(w)}\right) = O\left(\frac{1}{n^{2\cdot\mathrm{cl}(w)-1}}\right). \tag{1.4}$$

Remark 1.9. Recall that the Euler characteristic of an orientable compact surface of genus-g and ℓ boundary components is $2-2g-\ell$. Thus, Theorem 1.7 yields that the Laurent series of $\mathcal{T}r_{w_1,\ldots,w_\ell}(n)$ is supported on odd (respectively even) powers of n if ℓ is odd (respectively even). This is a nice interpretation of a fact that can also be derived directly from analysis involving Weingarten calculus.

Algebraic interpretation

The connection between the commutator length of a word w and $\chi_{\max}(w)$ led to the algebraic interpretation (1.4) in Corollary 1.8. This algebraic perspective also gives an interesting interpretation to our main theorem. Because a connected surface Σ is a $K(\pi_1(\Sigma),1)$-space, the Dehn-Nielsen-Baer Theorem states there is a natural isomorphism between $\mathrm{MCG}(\Sigma)$ and a certain subgroup of $\mathrm{Aut}(\pi_1(\Sigma))$ (see, for example, [FM12, Chapter 8] for its version for closed surfaces, and [MP15, Thm 2.4]). For example, if $\Sigma_{g,1}$ is a connected genus g surface with one boundary component, then $\pi_1(\Sigma) \cong \mathbf{F}_{2g} = \mathbf{F}(a_1,b_1,\ldots,a_g,b_g)$, and $\mathrm{MCG}(\Sigma)$ is isomorphic to the stabilizer of $[a_1,b_1]\cdots[a_g,b_g]$ in $\mathrm{Aut}(\mathbf{F}_{2g})$ – stabilizing this element reflects the fact that mapping classes in $\mathrm{MCG}(\Sigma)$ fix the boundary of $\Sigma_{g,1}$.

Along these lines, the set $\mathsf{Surfaces}(w)$ can be interpreted as equivalence classes of solutions to the equations

$$[u_1,v_1]\cdots[u_g,v_g] = w \tag{1.5}$$

with $u_i,v_i \in \mathbf{F}_r$ and varying g, where the equivalence relation is given by the action of the stabilizer $\mathrm{Aut}(\mathbf{F}_{2g})_{[a_1,b_1]\cdots[a_g,b_g]}$. In particular, the pairs $[(\Sigma,f)] \in \mathsf{Surfaces}(w)$ with maximal $\chi(\Sigma)$ correspond to equivalence classes of solutions to (1.5) with $g = \mathrm{cl}(w)$ minimal. Often, these solutions have trivial stabilizers, in which case $\chi^{(2)}(\mathrm{MCG}(f)) = 1$. For example, the stabilizer is trivial if the solutions consist of $2g$ free words, or, equivalently, if $f\colon \Sigma \to \bigvee^r S^1$ is π_1-injective – see Lemma 5.1. Thus, one could say

> *"The leading coefficient of $\mathcal{T}r_w(n)$ counts the number of equivalence classes of solutions to (1.5) with $g = \mathrm{cl}(w)$, up to corrections for the existence of non-trivial stabilizers."*

[5] A more general concept of the commutator length was introduced by Calegari (e.g. [Cal09a, Definition 2.71]), and applies to finite sets of words $w_1,\ldots,w_\ell$. This number can be related, under certain restrictions, to $\chi_{\max}(w_1,\ldots,w_\ell)$, in a similar fashion to the $\ell = 1$ case.

Examples

Let us now illustrate Theorem 1.7 and Corollary 1.8 on some of the examples from Table 1. The techniques by which we obtain some of the details in the following cases are explained throughout the paper, especially in Section 5.2.

- The commutator length of $w=[x,y]$ is obviously one, and there is a single equivalence class of solutions to the equation $[u,v]=w$, or, equivalently, a single element $[(\Sigma,f)]\in \mathsf{Surfaces}(w)$ with $\chi(\Sigma)=\chi_{\max}(w)=-1$. The stabilizer $\mathrm{MCG}(f)$ is trivial and so the first term in the Laurent expansion of $\mathcal{T}r_{[x,y]}(n)$ is $\frac{1}{n}$. Every other element $[(\Sigma,f)]\in\mathsf{Surfaces}(w)$ has $\chi(\Sigma)\le -3$ and $\chi^{(2)}(\mathrm{MCG}(f))=0$.

- We have $\mathrm{cl}\left(\left[x^3,y\right]\right)=1$. There are exactly three in-equivalent solutions to $[u,v]=w$: $\left[x^3,y\right]$, $\left[x^3,yx\right]$ and $\left[x^3,yx^2\right]$. In contrast, the solution $\left[x^3,yx^3\right]$ is equivalent to $\left[x^3,y\right]$ because the automorphism of $\mathbf{F}_2=\mathbf{F}(a,b)$ fixing a and mapping $b\to ba$, stabilizes $[a,b]$. In this case, all three solutions have trivial stabilizers, hence the leading term of $\mathcal{T}r_{[x^3,y]}$ is $\frac{3}{n}$. It seems like there are no other elements of $\mathsf{Surfaces}\left(\left[x^3,y\right]\right)$ with non-vanishing $\chi^{(2)}(\mathrm{MCG}(f))$ (at least there are none with $\chi(\Sigma)\ge -7$).

- In general, if $\mathrm{cl}(w)=1$, then every solution to $[u,v]=w$ has trivial stabilizer, because u and v are necessarily free words inside $\mathbf{F}_r$ (namely, $\langle u,v\rangle\cong\mathbf{F}_2$). Thus, for words of commutator length one we have $\mathcal{T}r_w(n)=\frac{K}{n}+O\left(\frac{1}{n^3}\right)$, where K is the number of equivalence classes of ways to write w as a commutator. Likewise, every solution to (1.5) with $\langle u_1,v_1,\ldots,u_g,v_g\rangle\cong\mathbf{F}_{2g}$ has a trivial stabilizer – see Lemma 5.1.

- For $w=[x,y]^2$ we have $\mathrm{cl}(w)=2$. There is a single equivalence class of solutions to (1.5) with $g=2$, and $\mathrm{MCG}(f)\cong\mathbf{F}_5$. As a bouquet of five circles is a $K(\mathbf{F}_5,1)$-space, we have $\chi(\mathbf{F}_5)=-4$. This explains the leading term of $\mathcal{T}r_{[x,y]^2}$.

- The somewhat surprising fact that $\mathrm{cl}\left([x,y]^3\right)=2$ was pointed out in [Cul81]. (Interestingly, Culler shows in that paper that $\mathrm{cl}([x,y]^n)=\left\lfloor\frac{n}{2}\right\rfloor+1$.) For example, $[x,y]^3=\left[xyx^{-1},y^{-1}xyx^{-2}\right]\left[y^{-1}xy,y^2\right]$. There are nine in-equivalent solutions in this case, each with a trivial stabilizer. This explain the leading term $\frac{9}{n^3}$. There is a single pair $[(\Sigma,f)]\in\mathsf{Surfaces}\left([x,y]^3\right)$ with $\chi(\Sigma)=-5$ and $\chi^{(2)}(\mathrm{MCG}(f))\ne 0$. The stabilizer in this single pair satisfies $\chi^{(2)}(\mathrm{MCG}(f))=81$. This explain the term $\frac{81}{n^5}$.

- The word $w=[x,y][x,z]$ has $\mathrm{cl}(w)=2$ and admits a single solution to (1.5) with $g=2$. The stabilizer of this solution is isomorphic to $\mathbb{Z}$ and $\chi^{(2)}(\mathbb{Z})=\chi(\mathbb{Z})=0$. Note that this explains why the coefficient of n^{-3} vanishes, but not why $\mathcal{T}r_w(n)=0$.

- The word $w=[x,y][x,z][x,t]$ has $\mathrm{cl}(w)=3$ and admits a single solution to (1.5) with $g=3$. The stabilizer of this solution is isomorphic to $\mathbb{Z}\times\mathbf{F}_2$ and $\chi^{(2)}(\mathbb{Z}\times\mathbf{F}_2)=\chi(\mathbb{Z}\times\mathbf{F}_2)=0$ (consult [MP15, Page 59] for more details).

- For $w_1=x^2y^2$ and $w_2=xy^{-3}x^{-3}y$ we have $\chi_{\max}(w_1,w_2)=-2$. There are four $[(\Sigma,f)]\in\mathsf{Surfaces}(w_1,w_2)$ with $\chi(\Sigma)=-2$, each with a trivial stabilizer, hence the leading term $\frac{4}{n^2}$. All $\chi(\Sigma)=-4$ solutions have $\chi^{(2)}(\mathrm{MCG}(f))=0$, while there is a single solution with non-vanishing contribution and $\chi(\Sigma)=-6$, for which $\chi^{(2)}(\mathrm{MCG}(f))=-16$.

- For every $w\ne 1$, $\chi_{\max}\left(w,w^{-1}\right)=0$ because there is an obvious annulus in $\mathsf{Surfaces}\left(w,w^{-1}\right)$. In both examples of this sort in Table 1, there is a single such annulus, and with a trivial stabilizer, hence the leading term 1. In both cases there are no other incompressible pairs in $\mathsf{Surfaces}\left(w, w^{-1}\right)$, but while for $w = x^2yxy^{-1}$, it seems that every compressible pair has vanishing contribution to $\mathcal{T}r_{w,w^{-1}}(n)$, for $w = x^2y^2xy^{-1}$ there is a compressible pair $[(\Sigma, f)]$ with $\chi(\Sigma) = -4$ and $\chi^{(2)}(\mathrm{MCG}(f)) = -4$.

Compressible vs. incompressible pairs $[(\Sigma, f)] \in \mathsf{Surfaces}(w_1, \ldots, w_\ell)$

The difference between compressible and incompressible pairs $[(\Sigma, f)] \in \mathsf{Surfaces}(w_1, \ldots, w_\ell)$ is already apparent from the fact that Theorem 1.6, or at least its proof, apply only to the incompressible case. The crucial property of incompressible pairs will be pointed out in Section 4.4 in the sequel of the paper. But there are some further differences we point out here.

First, there are finitely many incompressible elements in $\mathsf{Surfaces}(w_1, \ldots, w_\ell)$ – see Corollary 2.14. Because highest-Euler-characteristic elements are always incompressible, we deduce there are finitely many elements $[(\Sigma, f)] \in \mathsf{Surfaces}(w_1, \ldots, w_\ell)$ with $\chi(\Sigma) = \chi_{\max}(w_1, \ldots, w_\ell)$. In addition, as the examples above illustrate, the stabilizer $\mathrm{MCG}(f)$ of an incompressible solution is often trivial.

In contrast, there are infinitely many compressible elements in $\mathsf{Surfaces}(w_1, \ldots, w_\ell)$. In fact, there are often even infinitely many compressible elements $[(\Sigma, f)]$ with $\chi(\Sigma) = \chi_0$ for a given non-maximal χ_0, namely, for $\chi_0 = \chi_{\max}(w_1, \ldots, w_\ell) - 2k$ with $k \in \mathbb{Z}_{\geq 1}$, although, as stated in Theorem 1.7, almost all of them have zero contribution to $\mathcal{T}r_{w_1,\ldots,w_\ell}(n)$. Moreover, the stabilizer of a compressible pair is never trivial: a Dehn twist along a null-curve is a non-trivial element in the stabilizer.

1.1 Related lines of work

The evaluation of the integrals in (1.1) is a fundamental issue relating to several different areas of mathematics.

I. Matrix integrals in Gaussian models The connection between the enumeration of graphs on surfaces and matrix integrals in the classical GUE, GOE and GSE models was first established by 't Hooft [tH74] and later rediscovered by Harer and Zagier [HZ86]. For example, let $\mathrm{GUE}(n)$ denote the probability space of $n \times n$ Hermitian complex matrices endowed with complex Gaussian measure on each entry, where the (i, j) entry is independent of all other entries except for (j, i). The following equation [LZ04, Proposition 3.3.1] illustrates this connection:

$$\mathbb{E}_{H \in \mathrm{GUE}(n)}\left[(\mathrm{tr}H)^{\alpha_1}\left(\mathrm{tr}H^2\right)^{\alpha_2}\cdots\left(\mathrm{tr}H^k\right)^{\alpha_k}\right] = \sum_{\sigma} n^{F(\sigma)}. \tag{1.6}$$

The summation on the right hand side is over ribbon graphs (also known as fat-graphs) with α_i vertices of degree i for $i = 1, \ldots, k$, where σ is a perfect matching of the half-edges emanating from these vertices. The exponent $F(\sigma)$ is the number of faces in the embedding of the resulting ribbon graph on the surface of smallest possible genus. We stress that (1.6) is only an illustration of the theory, and there are many generalizations (e.g., for integrals over tuples of independent Hermitian matrices) and deep applications. For an excellent presentation of this theory, we refer the reader to [LZ04, Chapter 3].

There are many similarities between this by-now classical theory and the theory we develop in the current paper. For example, apart from the natural emergence of surfaces, the combinatorial formulas for $\mathcal{T}r_{w_1,\ldots,w_\ell}(n)$ we develop in Section 2 also involve a summation over perfect matchings. In addition, these matrix integrals over GUE were used, inter alia, to compute the Euler characteristic of the mapping class group of closed surfaces with punctures [HZ86, Pen88]. In fact, these Euler characteristics appear as coefficients in certain generating functions for integrals as in (1.6) (e.g., [Pen88, Theorem 1.1 and Corollary 3.1]).

There are also substantial differences. Among others, $\mathcal{U}(n)$ being a group endowed with Haar measure means that integrals as in (1.6) over a single Haar-random element can be completely computed using theoretical properties of the Haar measure, as was done in [DS94], and the computation becomes more interesting when multiple random elements are involved. It also means that word-measure on $\mathcal{U}(n)$ have nice properties, such as being $\mathrm{Aut}(\mathbf{F}_r)$-invariant, as explained in the following paragraph. In addition, the crucial role played here by maps from the surfaces to the bouquet of circles is completely absent in the classical theory. Another difference is that the summation in the right hand side of (1.6) is finite, with the exponents increasing as the Euler characteristic of the surface decreases. The best analogue in the current paper (2.10) involves an infinite summation with exponents decreasing together with the Euler characteristic of the surfaces. Finally, there is also a big difference in the role played by Euler characteristics of (subgroups of) the mapping class groups of surfaces.

II. Word measures on groups The same way $w \in \mathbf{F}_r$ induces a measure on $\mathcal{U}(n)$, it also induces a probability measure on any compact group (consult [HLS15] for recent results and references concerning the image of the word map $w\colon G^r \to G$ on compact Lie groups including $\mathcal{U}(n)$). By showing that Nielsen moves on w do not affect the resulting word measure, it is easy to see that two words in the same $\mathrm{Aut}(\mathbf{F}_r)$-orbit in $\mathbf{F}_r$ induce the same measure on every compact group (see [MP15, Section 2.2] for a proof). But is this the only reason for two words to have such a strong connection? A version of the following conjecture appears, for example, in [AV11, Question 2.2] and in [Sha13, Conjecture 4.2].

Conjecture 1.10. *If two words $w_1, w_2 \in \mathbf{F}_r$ induce the same measure on every compact group, then there exists $\phi \in \mathrm{Aut}(\mathbf{F}_r)$ with $w_2 = \phi(w_1)$.*

A special case of this conjecture deals with the $\mathrm{Aut}(\mathbf{F}_r)$-orbit of the single-letter word x_1, namely, with the set of primitive words. Several researchers have asked whether words inducing the Haar measure on every compact group are necessarily primitive. This was settled to the affirmative in [PP15, Theorem 1.1] using word measures on symmetric groups. In subsequent work [MP19b], we use the results in this paper and, mainly, Corollary 1.8, to prove that if a word w induces the same measure as $u_g = [x_1, y_1] \cdots [x_g, y_g]$ on every compact group then $w = \phi(u_g)$ for some $\phi \in \mathrm{Aut}(\mathbf{F}_r)$.

Short of proving Conjecture 1.10, one could hope to collect as many invariants of words as possible that can be determined by word measures induced on groups. For example, $\mathrm{cl}(w)$, the commutator length of a word, and more generally, $\chi_{\max}(w_1, \ldots, w_\ell)$, the highest possible Euler characteristic of a surface in $\mathsf{Surfaces}(w_1, \ldots, w_\ell)$, play an important role in our results. However, because the coefficient of $n^{\chi_{\max}(w_1,\ldots,w_\ell)}$ in $\mathcal{T}r_{w_1,\ldots,w_\ell}(n)$ occasionally vanishes, it is not clear whether $\mathrm{cl}(w)$ or $\chi_{\max}(w_1, \ldots, w_\ell)$ are determined by word measures on $\mathcal{U}(n)$.

In contrast, the measures do determine a related number, the *stable commutator length* of w. This algebraic quantity is defined by

$$\mathrm{scl}(w) \equiv \lim_{m\to\infty} \frac{\mathrm{cl}(w^m)}{m}. \tag{1.7}$$

(There is an analogous definition for finite sets of words.) There is a deep theory behind this invariant, and for background we refer to the short survey [Cal08] and long one [Cal09a] by Calegari. Relying on the rationality result of Calegari [Cal09b] that shows, in particular, that scl takes on rational values in $\mathbf{F}_r$, we are able to show the following:

Corollary 1.11. *The stable commutator length of a word $w \in [\mathbf{F}_r, \mathbf{F}_r]$ can be determined by the measures it induces on unitary groups in the following way:*

$$\mathrm{scl}(w) = \inf_{\ell>0;\, j_1,\ldots,j_\ell>0} \frac{-\lim_{n\to\infty} \log_n \left|\mathcal{T}r_{w^{j_1},\ldots,w^{j_\ell}}(n)\right|}{2(j_1 + \ldots + j_\ell)}. \tag{1.8}$$

A similar result is true for the stable commutator length of several words. We explain how Corollary 1.11 follows from Theorem 1.7 and Calegari's rationality theorem in Section 5.1.

Remark 1.12. In fact, with regards to Conjecture 1.10, word-measures on $\mathcal{U}(n)$ alone do not suffice and Conjecture 1.10 is not true if "every compact group" is replaced by "$\mathcal{U}(n)$ for all n". Indeed, for every $w \in \mathbf{F}_r$ and every n, the w-measure on $\mathcal{U}(n)$ is identical to the w^{-1}-measure. However, in general, w and w^{-1} belong to two different $\mathrm{Aut}(\mathbf{F}_r)$-orbits. See also Question 1 in Section 6.

III. Harmonic analysis on representation varieties. The integral in (1.1) can be viewed as an integral over the space of representations $\mathrm{Hom}(\mathbf{F}_r, \mathcal{U}(n))$ and in fact, as an integral over the representation variety

$$\mathrm{Rep}(\mathbf{F}_r, \mathcal{U}(n)) = \mathrm{Hom}(\mathbf{F}_r, \mathcal{U}(n)) / \mathcal{U}(n)$$

since the functions $\mathrm{tr} \circ w_i$ are invariant under $\mathcal{U}(n)$-conjugation, and so is the Haar measure. More generally, if Σ_g is the closed genus g surface, then the spaces $\mathrm{Rep}(\pi_1(\Sigma_g), \mathcal{U}(n))$ are of interest in geometry, via '*Higher Teichmüller theory*', dynamics as pioneered by Goldman [Gol97], and mathematical physics [Wit91]. For an overview see [Lab13]. For any closed curve on the surface, there is a natural function (*Wilson loop*) on the representation variety, given by the trace of the image of that curve in a given representation. It is natural to ask what is the integral of this function with respect to the volume form given by the Atiyah-Bott-Goldman symplectic structure on $\mathrm{Rep}(\pi_1(\Sigma_g), \mathcal{U}(n))$ [AB83, Gol84]. Our work answers this question for representations of free groups.

IV. Free probability theory. Voiculescu proved in [Voi91, Theorem 3.8] that for $w \neq 1$,

$$\mathcal{T}r_w(n) = o(1), \quad n \to \infty. \tag{1.9}$$

This is referred to the *asymptotic *-freeness* of the non-commutative independent random variables $(u_1, \ldots, u_r) \in \mathcal{U}(n)^r$, meaning that in the limit they can be modeled by the "Free Probability Theory" developed by Voiculescu (see, for example, the monograph [VDN92]). Such asymptotic freeness results are known for broad families of ensembles, including general Gaussian random matrices (due to Voiculescu in the same paper [Voi91, Theorem 2.2]). In later works (1.9) is strengthened to $\mathcal{T}r_w(n) = O\left(\frac{1}{n}\right)$ whenever $w \neq 1$ [MŚS07, Răd06]. Our work gives quantitative bounds on the decay rate of $\mathcal{T}r_w(n)$ (in many cases, from above and below) - see Corollary 1.8.

More generally, free probabilists are interested in the limit of $\mathcal{T}r_{w_1,\ldots,w_\ell}(n)$ as $n \to \infty$. This is given by the following corollary of our main result, which is essentially [MŚS07, Theorem 2] and [Răd06, Theorem 4.1]:

Corollary 1.13. *Let $w_1, \ldots, w_\ell \in \mathbf{F}_r$, each not equal to 1, and write $w_i = u_i^{d_i}$ where $u_i \in \mathbf{F}_r$ is a non-power and $d_i \geq 1$. Then the limit*

$$\lim_{n\to\infty} \mathcal{T}r_{w_1,\ldots,w_\ell}(n) \tag{1.10}$$

exists, and is equal to the number of ways to match $w_1, \ldots, w_\ell$ in pairs so that each word is conjugate to the inverse of its mate, times $\sqrt{\prod_{i=1}^{\ell} d_i}$.

Proof. As $w_1, \ldots, w_\ell \neq 1$, there are no surfaces of positive Euler characteristic in $\mathsf{Surfaces}(w_1, \ldots, w_\ell)$. The only possible surface Σ in this collection with $\chi(\Sigma) = 0$ is a disjoint union of annuli. The stabilizer $\mathrm{MCG}(f)$ is always trivial in this case, so the limit in (1.10) is equal to the number of such surfaces in $\mathsf{Surfaces}(w_1, \ldots, w_\ell)$. If w and w' are the words at the boundary of an annulus, then necessarily w' is conjugate to w^{-1}. Moreover, if $w = u^d$ with $u \in \mathbf{F}_r$ a non-power and $d \geq 1$, then the number of non-equivalent annuli in $\mathsf{Surfaces}(w, w^{-1})$ is exactly d. This yields the answer above. ∎

1.2 Paper organization

In Section 2 we show how surfaces emerge in the computation of $\mathcal{T}r_{w_1,\ldots,w_\ell}(n)$, present a formula for $\mathcal{T}r_{w_1,\ldots,w_\ell}(n)$ as a finite sum (Theorem 2.8) which yields Proposition 1.1, and then a second formula for $\mathcal{T}r_{w_1,\ldots,w_\ell}(n)$, this time as an infinite sum, but where the contribution of every surface is $\pm n^\alpha$ (Theorem 2.9). Section 2.5 then explains how every surface we constructed admits a natural map to the bouquet which makes it (a representative of) an element in $\mathsf{Surfaces}(w_1,\ldots,w_\ell)$. Thus, one can group together all the surfaces we constructed in the second formula (from Theorem 2.9) that belong to the same class $[(\Sigma,f)]\in\mathsf{Surfaces}(w_1,\ldots,w_\ell)$. Our main result then reduces to showing that the total contribution of this set of surfaces is equal to $\chi^{(2)}(\mathrm{MCG}(f))\cdot n^{\chi(\Sigma)}$, as stated in Theorem 1.7 -- this reduction is the content of Theorem 2.16.

In Sections 3 and 4 we fix $[(\Sigma,f)]\in\mathsf{Surfaces}(w_1,\ldots,w_\ell)$ and prove Theorem 2.16: in Section 3 we define the complex of transverse maps realizing (Σ,f), and prove it is a finite-dimensional contractible complex. In Section 4 we analyze the action of $\mathrm{MCG}(f)$ on this complex, show that the finite orbits of cells in this action are in one-to-one correspondence with the surfaces we constructed in Section 2, and finish the proof of Theorems 2.16, 1.4 and 1.7. In Sections 4.4 and 4.5 we discuss the difference between the compressible case and the incompressible one, and prove Theorem 1.6.

Section 5 contains three applications: in Section 5.1 we discuss stable commutator length and how it is determined by the w-measures on $\mathcal{U}(n)$, thus proving Corollary 1.11; in Section 5.2 we explain how our analysis yields a simple straight-forward algorithm to classify all incompressible solutions in $\mathsf{Surfaces}(w_1,\ldots,w_\ell)$, and, in particular, all solutions to the commutator equation (1.5) with $g=\mathrm{cl}(w)$; and in Section 5.3 we explain why $\mathrm{MCG}(f)$ has finite cohomological dimension. Section 6 contains some open questions.

Remark 1.14. The case where some of the words among $w_1,\ldots,w_\ell$ are trivial is not interesting in the point of view of estimating the integrals $\mathcal{T}r_{w_1,\ldots,w_\ell}(n)$: $\mathcal{T}r_{w_1,\ldots,w_{\ell-1},1}(n)=n\cdot\mathcal{T}r_{w_1,\ldots,w_{\ell-1}}(n)$. Yet, some of the results, such as Theorem 1.4, are interesting in this case too. Despite that, for the sake of simplicity, we assume throughout the rest of the paper that $w_i\neq 1$ for $i=1,\ldots,\ell$: this allows us to avoid extra case analysis at some points and shorten the arguments a bit. We stress, though, that all the results hold in the general case, and the proofs hold after, possibly, minor adaptations (with the one exception of Lemma 3.12 where, if one allows trivial words, the bound should be modified).

Remark 1.15. The unpublished manuscript [MP15] is based on an earlier stage of the current research. It contains some of the results of the current paper – mainly the results for incompressible maps – although with quite a different presentation of the proofs. Since writing [MP15], we have extended our results a great deal, and decided to rewrite everything in a whole new paper. To keep the current paper in manageable size, we include only ingredients that are necessary for proving and clarifying our results. Occasionally, we refer here to the more elaborated [MP15] for some background material, which is not used in the proofs.

Remark 1.16. A sequel paper [MP19a] shows how the ideas in the current paper can be extended and twisted to also deal with integrals over the orthogonal and compact symplectic groups.

Acknowledgments

We would like to thank Danny Calegari, Alexei Entin, Mark Feighn, Alex Gamburd, Yair Minsky, Mark Powell, Peter Sarnak, Zlil Sela, Karen Vogtmann, Alden Walker, Avi Wigderson, Qi You, and Ofer Zeitouni for valuable discussions about this work.

Part of this research was carried out during research visits of the first named author (Magee) to the Institute for Advanced Study in Princeton, and we would like to thank the I.A.S. for making these visits possible.

This is a pre-print of an article published in Inventiones Mathematicae. The final authenticated version is available online at: https://doi.org/10.1007/s00222-019-00891-4.

2 Combinatorial formulas for $\mathcal{T}r_{w_1,\ldots,w_\ell}(n)$ using surfaces

In this section we recall basic results about the Weingarten calculus for integrals over $\mathcal{U}(n)$, and derive formulas for $\mathcal{T}r_{w_1,\ldots,w_\ell}(n)$ which involve surfaces. But first, we explain why $\mathcal{T}r_{w_1,\ldots,w_\ell}(n)$ vanishes in the "non-balanced" case, where the total exponent of some letter is not zero.:

Claim 2.1. Let $w_1,\ldots,w_\ell \in \mathbf{F}_r$. If $w_1 w_2 \cdots w_\ell \notin [\mathbf{F}_r,\mathbf{F}_r]$ then

1. $\mathcal{T}r_{w_1,\ldots,w_\ell}(n) \equiv 0$.
2. The set $\mathsf{Surfaces}(w_1,\ldots,w_\ell)$ is empty.

Proof. (1) The assumption $w_1,\ldots,w_\ell \notin [\mathbf{F}_r,\mathbf{F}_r]$ is equivalent to that there is some $j \in [r]$ so that α_j, the sum of exponents of the letter x_j in $w_1,\ldots,w_\ell$, satisfies $\alpha_j \neq 0$. As the Haar measure of a compact group is invariant under left multiplication by any element, and the diagonal central matrix $e^{i\theta} I_n$ is in $\mathcal{U}(n)$ for $\theta \in [0,2\pi]$, we obtain

$$\begin{aligned}
&\mathcal{T}r_{w_1,\ldots,w_\ell}(n) = \\
&= \mathbb{E}_{A_1,\ldots,A_r\in\mathcal{U}(n)}\left[\mathrm{tr}\left(w_1\left(A_1,\ldots,A_j,\ldots,A_r\right)\right)\cdots\mathrm{tr}\left(w_\ell\left(A_1,\ldots,A_j,\ldots,A_r\right)\right)\right] \\
&= \mathbb{E}_{A_1,\ldots,A_r\in\mathcal{U}(n)}\left[\mathrm{tr}\left(w_1\left(A_1,\ldots,e^{i\theta}A_j,\ldots,A_r\right)\right)\cdots\mathrm{tr}\left(w_\ell\left(A_1,\ldots,e^{i\theta}A_j,\ldots,A_r\right)\right)\right] \\
&= e^{i\theta\alpha_j}\cdot\mathcal{T}r_{w_1,\ldots,w_\ell}(n).
\end{aligned}$$

The first statement follows as this equality holds for every $\theta \in [0,2\pi]$.

(2) The second statement follows from the fact that in every connected, orientable, compact surface Σ with boundary, the product in $\pi_1(\Sigma)$ of loops around the boundary components belongs to $[\pi_1(\Sigma),\pi_1(\Sigma)]$. □

2.1 Weingarten function and integrals over $\mathcal{U}(n)$

The "Weingarten calculus" for computing integrals over unitary groups with respect to the Haar measure was developed in a series of papers, most notably [Wei78, Xu97, Col03, CŚ06]. It is based on the Schur-Weyl duality (see Remark 2.6 below), and allows the computation of integrals over the entries of unitary matrices and their complex conjugates, as depicted in Theorem 2.5 below. This computation is given in terms of the Weingarten function, which we now describe.

Let $\mathbb{Q}(n)$ denote the field of rational functions with rational coefficients in the variable n. Let S_L denote the symmetric group on L elements. For every $L \in \mathbb{Z}_{\geq 1}$, the Weingarten function Wg_L S_L
maps S_L to $\mathbb{Q}(n)$. We think of such functions as elements of the group ring $\mathbb{Q}(n)[S_L]$.

Definition 2.2. The **Weingarten function** $\mathrm{Wg}_L : S_L \to \mathbb{Q}(n)$ is the inverse, in the group ring Wg_L
$\mathbb{Q}(n)[S_L]$, of the function $\sigma \mapsto n^{\#\mathrm{cycles}(\sigma)}$.

That the function $\sigma \mapsto n^{\#\mathrm{cycles}(\sigma)}$ is invertible for every L follows from [CŚ06, Proposition 2.3] and the discussion following it. In particular, $\mathrm{Wg}_L(\sigma)$ is in $\mathbb{Q}(n)$ for every $\sigma \in S_L$. Clearly, Wg_L is constant on conjugacy classes. For example, for $L=2$, the inverse of $\left(n^2\cdot\mathrm{id} + n\cdot(12)\right) \in \mathbb{Q}(n)[S_2]$ is $\left(\frac{1}{n^2-1}\cdot\mathrm{id} - \frac{1}{n(n^2-1)}\cdot(12)\right)$, so $\mathrm{Wg}_2(\mathrm{id}) = \frac{1}{n^2-1}$ while $\mathrm{Wg}_2((12)) = \frac{-1}{n(n^2-1)}$. The values of Wg_3 are

$$\mathrm{id} \mapsto \frac{n^2-2}{n\left(n^2-1\right)\left(n^2-4\right)} \qquad (12) \mapsto \frac{-1}{\left(n^2-1\right)\left(n^2-4\right)} \qquad (123) \mapsto \frac{2}{n\left(n^2-1\right)\left(n^2-4\right)}.$$

Collins and Śniady provide an explicit formula for Wg_L in terms of the irreducible characters of S_L and Schur polynomials [CŚ06, Equation (13)]:

$$\mathrm{Wg}_L(\sigma) = \frac{1}{(L!)^2} \sum_{\lambda \vdash L} \frac{\chi_\lambda(e)^2}{d_\lambda(n)} \chi_\lambda(\sigma),$$

where λ runs over all partitions of L, χ_λ is the character of S_L corresponding to λ, and $d_\lambda(n)$ is the number of semistandard Young tableaux with shape λ, filled with numbers from $[n]$. A well known formula for $d_\lambda(n)$ is $d_\lambda(n) = \frac{\chi_\lambda(e)}{L!} \prod_{(i,j)\in\lambda} (n+j-i)$, where (i,j) are the coordinates of cells in the Young diagram with shape λ (e.g. [Ful97, Section 4.3, Equation (9)]). Thus,

Corollary 2.3. *For $\sigma \in S_L$, $\mathrm{Wg}_L(\sigma)$ may have poles only at integers n with $-L < n < L$.*

Below we use the following properties of the Weingarten function. The standard norm of $\rho \in S_L$, denoted $\|\rho\|$, is the shortest length of a product of transpositions giving ρ, and is equal to $L - \#\mathrm{cycles}(\rho)$.

Theorem 2.4. *Let $\pi \in S_L$ be a permutation.*

1. *[CŚ06, Corollary 2.7] Leading term:*

$$\mathrm{Wg}_L(\pi) = \frac{\mathrm{Möb}(\pi)}{n^{L+\|\pi\|}} + O\left(\frac{1}{n^{L+\|\pi\|+2}}\right), \tag{2.1}$$

where Möb (σ)

$$\mathrm{Möb}(\pi) = \mathrm{sgn}(\pi) \prod_{i=1}^{k} c_{|C_i|-1}, \tag{2.2}$$

with[6] *$C_1, \ldots, C_k$ the cycles composing π, and $c_m = \frac{(2m)!}{m!(m+1)!}$ being the m-th Catalan number.*

2. *[Col03, Theorem 2.2] Asymptotic expansion:*

$$\mathrm{Wg}_L(\pi) = \frac{1}{n^L} \sum_{k \in \mathbb{Z}_{\geq 0}} \sum_{\substack{\rho_1,\ldots,\rho_k \in S_L \setminus \{\mathrm{id}\} \\ \text{s.t. } \rho_1 \cdots \rho_k = \pi}} \frac{(-1)^k}{n^{\|\rho_1\|+\ldots+\|\rho_k\|}}. \tag{2.3}$$

(In (2.3), when $\pi = \mathrm{id}$, there is a term $\frac{1}{n^L}$ coming from $k = 0$.)

The Weingarten function is used in the following formula of Collins and Śniady, which evaluates integrals of monomials in the entries $A_{i,j}$ and their conjugates $\overline{A_{i,j}}$ of a Haar-random unitary matrix $A \in \mathcal{U}(n)$. As in the proof of Claim 2.1, this integral vanishes whenever the monomial is not balanced, namely whenever the number of $A_{i,j}$'s is different from the number of $\overline{A_{i,j}}$'s.

Theorem 2.5. *[CŚ06, Proposition 2.5] Let L be a positive integer and $(i_1, \ldots, i_L)$, $(j_1, \ldots, j_L)$, $(i'_1, \ldots, i'_L)$ and $(j'_1, \ldots, j'_L)$ be L-tuples of positive integers. Then for every n for which the expression*

$$\mathbb{E}_{A \in \mathcal{U}(n)} \left[A_{i_1,j_1} A_{i_2,j_2} \ldots A_{i_L,j_L} \overline{A_{i'_1,j'_1} A_{i'_2,j'_2}} \ldots \overline{A_{i'_L,j'_L}} \right] \tag{2.4}$$

makes sense, namely, for $n \geq \max\{i_1, \ldots, i_L, j_1, \ldots, j_L, i'_1, \ldots, i'_L, j'_1, \ldots, j'_L\}$, (2.4) is equal to the evaluation of n in a rational function, which is given by

$$\sum_{\sigma,\tau \in S_L} \delta_{i_1 i'_{\sigma(1)}} \ldots \delta_{i_L i'_{\sigma(L)}} \delta_{j_1 j'_{\tau(1)}} \ldots \delta_{j_L j'_{\tau(L)}} \mathrm{Wg}_L\left(\sigma^{-1}\tau\right). \tag{2.5}$$

[6] The function Möb is the Möbius function on a natural poset structure on S_L – see, for instance, [NS06, Lectures 10 and 23].

Put differently, the rational function is given by $\sum_{\sigma,\tau} \mathrm{Wg}_L\left(\sigma^{-1}\tau\right)$, where σ runs over all rearrangements of $(i'_1,\ldots,i'_L)$ which make it identical to $(i_1,\ldots,i_L)$, and τ runs over all rearrangements of $(j'_1,\ldots,j'_L)$ which make it identical to $(j_1,\ldots,j_L)$. In particular, the possible poles of the Weingarten function at n, for every $n \geq \max\{i_1,\ldots,j'_L\}$, are guaranteed to cancel out in this summation (see the example following Proposition 2.5 in [CŚ06]).

Remark 2.6. The basis for the Weingarten calculus is the Schur-Weyl duality for $\mathcal{U}(n)$. One version of this duality is the following: let $V = \mathbb{C}^n$. A unitary matrix $A \in \mathcal{U}(n)$ acts on the space of functionals $W = \left(V^{\otimes L} \otimes (V^*)^{\otimes L}\right)^*$ by

$$(A\theta)(v_1 \otimes \ldots \otimes v_L \otimes \varphi_1 \otimes \ldots \otimes \varphi_L) = \theta\left(A^{-1}v_1 \otimes \ldots \otimes A^{-1}v_L \otimes A^{-1}\varphi_1 \otimes \ldots \otimes A^{-1}\varphi_L\right),$$

where we think of $\varphi \in V^*$ as a column vector in $\mathbb{C}^n$ whose value on $v \in V$ is $\varphi^* v \in \mathbb{C}$. Every permutation $\sigma \in S_L$ yields a functional in W defined by:

$$\theta_\sigma(v_1 \otimes \ldots \otimes v_q \otimes \varphi_1 \otimes \ldots \otimes \varphi_q) = \left(\varphi_1^* v_{\sigma^{-1}(1)}\right)\cdots\left(\varphi_q^* v_{\sigma^{-1}(q)}\right).$$

The Schur-Weyl duality says that this embedding of $\mathbb{C}[S_L]$ in W is precisely the set of $\mathcal{U}(n)$-invariant functionals in W. The family of integrals in (2.4) can be presented as a single functional on $V^{\otimes L} \otimes (V^*)^{\otimes L} \otimes V^{\otimes L} \otimes (V^*)^{\otimes L}$, which is $\mathcal{U}(n) \times \mathcal{U}(n)$-invariant because the Haar measure is both left- and right-invariant. This roughly explains why one can expect a result of the type of Theorem 2.5.

2.2 Surfaces from matchings of letters

Based on Theorem 2.5 we show that $\mathcal{T}r_{w_1,\ldots,w_\ell}(n)$ is a rational expression in n, and give concrete formulas which involve surfaces. These surfaces are constructed from matchings of the letters in $w_1,\ldots,w_\ell$, and we begin by describing this construction.

Recall that B denotes a fixed basis for $\mathbf{F}_r$. Following Claim 2.1, we assume that $w_1 \cdots w_\ell \in [\mathbf{F}_r, \mathbf{F}_r]$, namely, that for every letter $x \in B$, the total number of instances of x^{+1} in $w_1,\ldots,w_\ell$ is equal to that of x^{-1}, and we denote this number by $L_x \in \mathbb{Z}_{\geq 0}$. In particular, $|w_1| + \ldots + |w_\ell| = 2\sum_{x\in B} L_x$. We also denote by $\mathrm{MATCH}_x(w_1,\ldots,w_\ell)$ the set of bijections from the instances of x^{+1} to the instances of x^{-1}, so that $|\mathrm{MATCH}_x(w_1,\ldots,w_\ell)| = L_x!$. L_x MATCH_x

Let $\kappa = \{\kappa_x\}_{x\in B} \in (\mathbb{Z}_{\geq 0})^B$ be an assignment of a non-negative integer to every basis element. We denote by $\mathrm{MATCH}^\kappa(w_1,\ldots,w_\ell)$ the Cartesian product of sets of matchings, with $\kappa_x + 1$ copies of $\mathrm{MATCH}_x(w_1,\ldots,w_\ell)$ for every $x \in B$, namely, κ MATCH^κ

$$\mathrm{MATCH}^\kappa(w_1,\ldots,w_\ell) \stackrel{\mathrm{def}}{=} \prod_{x\in B} \mathrm{MATCH}_x(w_1,\ldots,w_\ell)^{\kappa_x+1}.$$

The following definition presents the construction of a surface from an element of $\mathrm{MATCH}^\kappa(w_1,\ldots,w_\ell)$. We use the notation $[k] \stackrel{\mathrm{def}}{=} \{0,1,\ldots,k\}$ for a non-negative integer k. $[k]$

Definition 2.7. Let $w_1,\ldots,w_\ell \in \mathbf{F}_r \setminus \{1\}$ be a balanced set of words, let $\kappa \in (\mathbb{Z}_{\geq 0})^B$ and let $\sigma \in \mathrm{MATCH}^\kappa(w_1,\ldots,w_\ell)$ be a tuple of matchings. We denote by $\sigma_{x,0},\ldots,\sigma_{x,\kappa_x}$ the $\kappa_x + 1$ matchings from $\mathrm{MATCH}_x(w_1,\ldots,w_\ell)$ in σ. From this data we construct a surface Σ_σ as a CW-complex as follows: Σ_σ

- For $1 \neq w \in \mathbf{F}_r$ define $S^1(w)$ to be an oriented 1-sphere S^1 with additional marked points as follows: there are[7] $|w|$ points marked o, which we call o-points. These points cut the 1-sphere into $|w|$ intervals, which are in bijection with the letters of w, in the suitable cyclic $S^1(w)$ o-point

[7] We use $|w|$ to denote the number of letters in w.

order. For every letter of w, if the letter is $x^{\pm 1}$, we mark additional $\kappa_x + 1$ points on the interval corresponding to that letter. These marked points are labeled $(x, 0), \ldots, (x, \kappa_x)$ and are ordered according to the orientation of $S^1(w)$ if the letter is x^{+1}, or in reverse orientation if the letter is x^{-1}. We call a point labeled (x, j) for some $x \in B$ and $j \in [\kappa_x]$ an (x, j)-point or a z-point if the exact x and j do not matter. (x, j)-point z-point

- The one-dimensional skeleton of Σ_σ consists of $S^1(w_1), \ldots, S^1(w_\ell)$, together with additional $\sum_{x \in B} L_x(\kappa_x + 1)$ edges (1-cells), referred to as *matching-edges*: for every $x \in B$ and $j \in [\kappa_x]$, introduce L_x edges describing the matching $\sigma_{x,j}$. Namely, for every x^{+1}-letter λ of $w_1, \ldots, w_\ell$, introduce an edge between the (x, j)-point on the interval corresponding to λ and the (x, j)-point on the interval corresponding to the x^{-1}-letter $\sigma_{x,j}(\lambda)$. This is illustrated in the left part of Figure 2.1.

- Finally, 2-cells are attached as follows: consider cycles in the 1-skeleton which are obtained by starting at some marked point in $S^1(w_i)$ for some $i = 1, \ldots, \ell$, moving orientably along $S^1(w_i)$ until the next z-point, then following the matching-edge emanating from this z-point and arriving at some z-point in $S^1(w_{i'})$ for some i', then moving orientably along $S^1(w_{i'})$ until the next z-point, continuing along the matching-edge and so on until a cycle has been completed. A 2-cell (a disc) is glued along every such cycle.

- From the construction of Σ_σ, it is clear it is a surface, with boundary $S^1(w_1) \sqcup \ldots \sqcup S^1(w_\ell)$ and with orientation prescribed from the boundary. Moreover, every 2-cell D belongs to exactly one of the following categories:

  - Either there is an o-point at every component of $\partial D \cap \partial \Sigma_\sigma$, in which case we call D an o-disc, o-disc
  - or, ∂D contains no o-points, in which case we call D a z-disc. In this case, there are some $x \in B$ with $\kappa_x \geq 1$ and $j \in [\kappa_x - 1]$ such that the marked points in ∂D are exactly of two types: (x, j)-points and $(x, j+1)$-points. In this case we call the z-disc D also an (x, j)-disc. See Figure 2.1. z-disc (x, j)-disc

- Let $\chi(\sigma)$ denote the Euler characteristic of this surface, namely $\chi(\sigma) \stackrel{\text{def}}{=} \chi(\Sigma_\sigma)$. $\chi(\sigma)$

2.3 A formula for $\mathcal{T}r_{w_1,\ldots,w_\ell}(n)$ as a rational expression

Our first formula for $\mathcal{T}r_{w_1,\ldots,w_\ell}(n)$ is a finite sum over pairs of matchings for every letters, namely over elements in $\mathrm{MATCH}^\kappa(w_1, \ldots, w_\ell)$ with $\kappa_x = 1$ for every $x \in B$. We denote this κ by $\kappa \equiv 1$. In particular, this formula proves Proposition 1.1.

Theorem 2.8 ($\mathcal{T}r_{w_1,\ldots,w_\ell}(n)$ as finite sum)**.** *Let $w_1, \ldots, w_\ell \in \mathbf{F}_r$ be a balanced set of words.*

1. *If[8] $n \geq L_x$ for every $x \in B$, then*

$$\mathcal{T}r_{w_1,\ldots,w_\ell}(n) = \sum_{\sigma \in \mathrm{MATCH}^{\kappa \equiv 1}} \left(\prod_{x \in B} \mathrm{Wg}_{L_x}\left(\sigma_{x,0}^{-1} \sigma_{x,1}\right) \right) \cdot n^{\#\{o\text{-discs in } \Sigma_\sigma\}} \tag{2.6}$$

   (here $\sigma_{x,0}^{-1}\sigma_{x,1}$ is a permutation of the x^{+1}-letters of $w_1, \ldots, w_\ell$).

2. *For $n \geq \max_{x \in B} L_x$, the function $\mathcal{T}r_{w_1,\ldots,w_\ell}(n)$ is a computable rational function in n.*

[8] Interestingly, very similar constraints on n appear in a formula giving the expected trace of w in r uniform $n \times n$ *permutation* matrices as a rational expression in n – see [Pud14, Section 5].

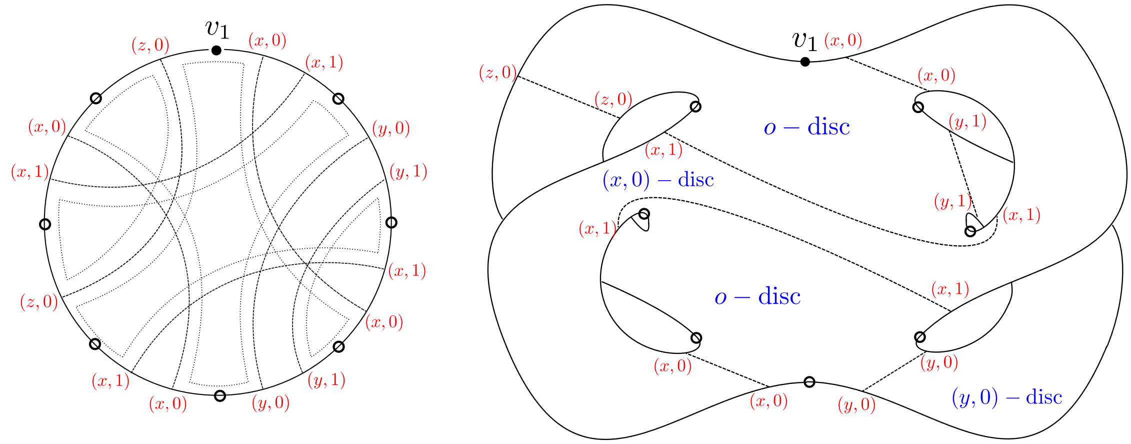

Figure 2.1: On the left is the 1-skeleton of Σ_σ for $w = [x,y][x,z] = x_1y_2X_3Y_4x_5z_6X_7Z_8$, with $\kappa_x = \kappa_y = 1$ and $\kappa_z = 0$ and with the matchings $\sigma_{x,0} = (x_1 \mapsto X_3;\ x_5 \mapsto X_7)$, $\sigma_{x,1} = (x_1 \mapsto X_7;\ x_5 \mapsto X_3)$, $\sigma_{y,0} = \sigma_{y,1} = (y_2 \mapsto Y_4)$ and $\sigma_{z,0} = (z_6 \mapsto Z_8)$. Dashed lines are matching-edges. The dotted lines trace the boundaries of the two o-discs to be glued in (see Definition 2.7). Two additional discs, a $(x,0)$-disc and a $(y,0)$-disc are glued along the other types of cycles one can follow (unmarked). The eight o-points are marked by V_1 and black circles. The resulting surface Σ_σ is on the right and is a genus-2 surface with one boundary component. In this case, $\chi(\sigma) = \chi(\Sigma_\sigma) = -3$.

3. *For $\sigma \in \mathrm{MATCH}^{\kappa\equiv 1}(w_1,\dots,w_\ell)$, let $\sigma_0 = (\sigma_{x,0})_{x\in B}$ and $\sigma_1 = (\sigma_{x,1})_{x\in B}$ denote two matchings of the "positive" letters of $w_1,\dots,w_\ell$ to the "negative" ones. Then the summand in (2.6) corresponding to σ is*
$$\text{Möb}\left(\sigma_0^{-1}\sigma_1\right)\cdot n^{\chi(\sigma)} + O\left(n^{\chi(\sigma)-2}\right). \tag{2.7}$$

Proof. Part 2 follows from (2.6) as every value of the Weingarten function is computable and in $\mathbb{Q}(n)$. We now prove part (1), which we do by way of an example. Let $w_1 = xyx^{-2}y$ and $w_2 = xy^{-2}$. Then,

$$\begin{aligned}
\mathcal{T}r_{w_1,w_2}(n) &= \mathbb{E}_{(A,B)\in\mathcal{U}(n)\times\mathcal{U}(n)}\left[\mathrm{tr}\left(ABA^{-2}B\right)\cdot\mathrm{tr}\left(AB^{-2}\right)\right]\\
&= \mathbb{E}_{(A,B)\in\mathcal{U}(n)\times\mathcal{U}(n)}\left[\left(\sum_{i,j,k,\ell,m\in[n]} A_{i,j}\cdot B_{j,k}\cdot A^{-1}_{k,\ell}\cdot A^{-1}_{\ell,m}\cdot B_{m,i}\right)\left(\sum_{I,J,K\in[n]} A_{I,J}\cdot B^{-1}_{J,K}\cdot B^{-1}_{K,I}\right)\right]\\
&= \sum_{i,j,k,\ell,m,I,J,K\in[n]} \mathbb{E}_{(A,B)\in\mathcal{U}(n)\times\mathcal{U}(n)}\left[A_{i,j}\cdot B_{j,k}\cdot\overline{A_{\ell,k}}\cdot\overline{A_{m,\ell}}\cdot B_{m,i}\cdot A_{I,J}\cdot\overline{B_{K,J}}\cdot\overline{B_{I,K}}\right]\\
&= \sum_{i,j,k,\ell,m,I,J,K\in[n]} \left(\mathbb{E}_{A\in\mathcal{U}(n)}\left[A_{i,j}\cdot A_{I,J}\cdot\overline{A_{\ell,k}}\cdot\overline{A_{m,\ell}}\right]\right)\cdot\left(\mathbb{E}_{B\in\mathcal{U}(n)}\left[B_{j,k}\cdot B_{m,i}\cdot\overline{B_{K,J}}\cdot\overline{B_{I,K}}\right]\right).
\end{aligned}\tag{2.8}$$

Note that there is a clear correspondence between the o-points in $S^1(w_1)$ and the indices i,j,k,ℓ,m and between the o-points in $S^1(w_2)$ and the indices I,J,K (see Figure 2.2).

Now we use Theorem 2.5 to replace each of the two integrals inside the sum by a summation over pairs of matchings. For the first integral we go over all bijections $\{i,I\}\xrightarrow{\sim}\{\ell,m\}$ and $\{j,J\}\xrightarrow{\sim}\{k,\ell\}$, and we think of them as elements $\sigma_{x,0},\sigma_{x,1}\in\mathrm{MATCH}_x(w_1,w_2)$ by thinking of a matching between two z-points as a matching of the adjacent o-points. For example, the $(x,0)$-point in the first letter of w_1 is adjacent to the o-point i, and the $(x,1)$-point in the same letter is adjacent to the o-point j. Similarly, we go over all bijections $\sigma_{y,0}$ and $\sigma_{y,1}$ for the second integral. Changing the order of

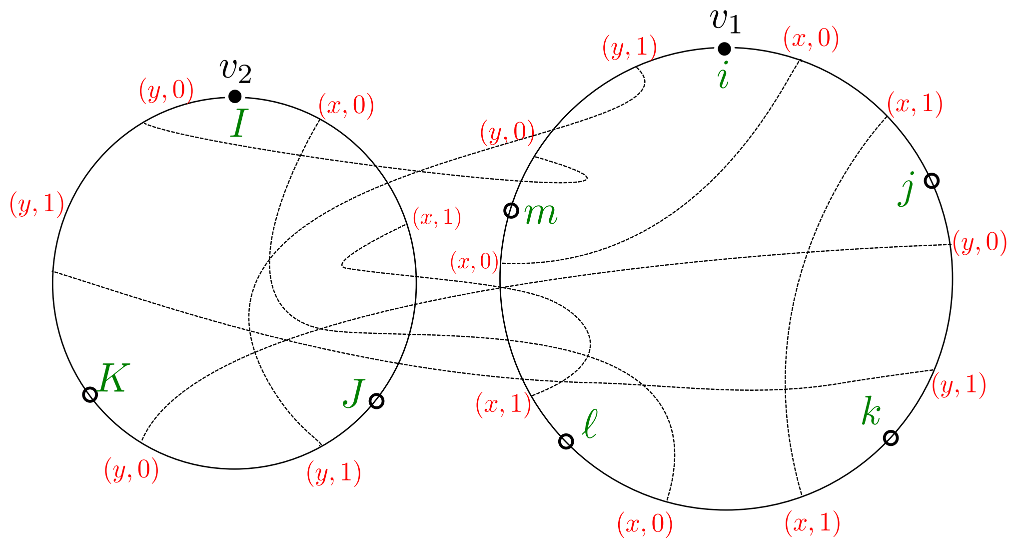

Figure 2.2: The 1-skeleton of the surface Σ_σ for the tuple of matchings $\sigma \in \mathrm{MATCH}^{\kappa\equiv 1}\left(xyx^{-2}y, xy^{-2}\right)$ specified in the proof of Theorem 2.8. The o-points are identified with the indices i, j, k, ℓ, m and I, J, K that appear in the computation of $\mathcal{T}r_{xyx^{-2}y,xy^{-2}}(n)$. The o-discs of Σ_σ (two of these in the current case) are in one-to-one correspondence with the blocks of indices determined by σ, and for every $x \in B$, the $(x,0)$-discs (one for each letter in the current case) are in one-to-one correspondence with the cycles of the permutation $\sigma_{x,0}^{-1}\sigma_{x,1}$.

summation, we sum first over $\sigma_{x,0}$, $\sigma_{x,1}$, $\sigma_{y,0}$ and $\sigma_{y,1}$, and only then over the indices $i, j, \ldots, K$. This turns (2.8) into a sum over $\mathrm{MATCH}^{\kappa\equiv 1}(w_1, w_2)$.

For every set of $\sigma \in \mathrm{MATCH}^{\kappa\equiv 1}(w_1, w_2)$, we only need to count the number of evaluations of $i, j, \ldots, L$ which "agree" with the permutations. For example, consider the case where

$$\begin{array}{cccc} \underline{\sigma_{x,0}} & \underline{\sigma_{x,1}} & \underline{\sigma_{y,0}} & \underline{\sigma_{y,1}} \\ i \mapsto m & j \mapsto k & j \mapsto K & k \mapsto K \\ I \mapsto \ell & J \mapsto \ell & m \mapsto I & i \mapsto J \end{array} \tag{2.9}$$

(these are the matchings described in Figure 2.2). Note that in this case, both the permutation $\sigma_{x,0}^{-1}\sigma_{x,1}$ and the permutation $\sigma_{y,0}^{-1}\sigma_{y,1}$ are a 2-cycle. Hence, by Theorem 2.5, the summand corresponding to these matchings is

$$\mathrm{Wg}_2((1\ 2)) \cdot \mathrm{Wg}_2((1\ 2)) \cdot \sum_{i,j,k,\ell,m,I,J,K \in [n]} \delta_{im}\delta_{I\ell}\delta_{jk}\delta_{J\ell}\delta_{jK}\delta_{mI}\delta_{kK}\delta_{iJ},$$

and the product inside the last sum is 1 (and not 0) if and only if $i = m = I = \ell = J$ and $j = k = K$. Here, two indices must have the same value if and only if they belong to the same o-disc in Σ_σ, hence there are exactly $n^{\#o-\text{discs in } \Sigma_\sigma}$ contributing values of the indices, each contributing 1 to the summation. For σ we defined in (2.9) this number is n^2, and the total contribution of this σ is, thus, $\mathrm{Wg}_2((1\ 2))^2 \cdot n^2 = \frac{1}{(n^2-1)^2}$. The total summation over all the 16 elements of $\mathrm{MATCH}^{\kappa\equiv 1}(w_1, w_2)$ is $\frac{1}{n^2-1}$. Since the same argument works for every $w_1, \ldots, w_\ell \in \mathbf{F}_r$, this proves part (1).

Recall that for $\pi \in S_L$, we have $\|\pi\| = L - \#\mathrm{cycles}(\pi)$. The number of cycles in the permutation

$\sigma_{x,0}^{-1}\sigma_{x,1} \in S_{L_x}$ is equal to the number of $(x,0)$-discs in Σ_σ. Hence, by Theorem 2.4(1),

$$\begin{aligned}\prod_{x\in B} \mathrm{Wg}_{L_x}\left(\sigma_{x,0}^{-1}\sigma_{x,1}\right) &= \prod_{x\in B}\left[\frac{\mathrm{M\ddot{o}b}\left(\sigma_{x,0}^{-1}\sigma_{x,1}\right)}{n^{L_x+\left\|\sigma_{x,0}^{-1}\sigma_{x,1}\right\|}} + O\left(\frac{1}{n^{L_x+\left\|\sigma_{x,0}^{-1}\sigma_{x,1}\right\|+2}}\right)\right] \\ &= \prod_{x\in B}\left[\frac{\mathrm{M\ddot{o}b}\left(\sigma_{x,0}^{-1}\sigma_{x,1}\right)}{n^{2L_x-\#\{(x,0)-\text{discs in }\Sigma_\sigma\}}} + O\left(\frac{1}{n^{2L_x-\#\{(x,0)-\text{discs in }\Sigma_\sigma\}+2}}\right)\right] \\ &= \frac{\mathrm{M\ddot{o}b}\left(\sigma_0^{-1}\sigma_1\right)}{n^{2L-\#\{z-\text{discs in }\Sigma_\sigma\}}} + O\left(\frac{1}{n^{2L-\#\{z-\text{discs in }\Sigma_\sigma\}-2}}\right),\end{aligned}$$

where $L = \sum_{x\in B} L_x$ is the total number of positive letters in $w_1,\ldots,w_\ell$. We are done proving part (3) as

$$\chi(\sigma) = \chi(\Sigma_\sigma) = -2L + \#\{\text{discs in }\Sigma_\sigma\}.$$

$\square$

2.4 A formula for $\mathcal{T}r_{w_1,\ldots,w_\ell}(n)$ as Laurent expansion

We now give an alternative formula for $\mathcal{T}r_{w_1,\ldots,w_\ell}(n)$, which also uses surfaces constructed from matchings of the letters of $w_1,\ldots,w_\ell$. The sum in (2.6) is finite, it proves the rationality of $\mathcal{T}r_{w_1,\ldots,w_\ell}(n)$, and allows a finite algorithm to compute it. The alternative formula we introduce next has the disadvantage that it is an infinite sum (unless $L_x \le 1$ for every $x \in B$). However, it has the advantage of simplifying greatly the contribution of every surface involved in the computation, as well as being an important step towards establishing Theorem 1.7. This formula is derived from (2.6) together the asymptotic expansion of the Weingarten function developed in [Col03] and depicted in Theorem 2.4(2) above[9].

The formula uses a restricted set of tuples of matchings which, for a given $\kappa \in (\mathbb{Z}_{\ge 0})^B$, we denote by $\overline{\mathrm{MATCH}}^\kappa(w_1,\ldots,w_\ell)$: this is the subset of $\mathrm{MATCH}^\kappa(w_1,\ldots,w_\ell)$ with the restriction that no two adjacent matchings are identical, namely, that $\sigma_{x,j} \ne \sigma_{x,j+1}$ for every $x\in B$ and $0 \le j \le \kappa_x - 1$. We also denote by $\overline{\mathrm{MATCH}}^*(w_1,\ldots,w_\ell)$ the union of restricted sets of matchings over all possible κ: $\overline{\mathrm{MATCH}}^\kappa$ $\overline{\mathrm{MATCH}}^*$

$$\overline{\mathrm{MATCH}}^*(w_1,\ldots,w_\ell) \stackrel{\text{def}}{=} \coprod_{\kappa\in\left(\mathbb{Z}_{\ge 0}\right)^B} \overline{\mathrm{MATCH}}^\kappa(w_1,\ldots,w_\ell),$$

and for $\sigma \in \overline{\mathrm{MATCH}}^*(w_1,\ldots,w_\ell)$ denote by $\kappa(\sigma)$ and $\kappa_x(\sigma)$ the corresponding values of κ and κ_x. Also, for $\kappa \in (\mathbb{Z}_{\ge 0})^B$ let $|\kappa| \stackrel{\text{def}}{=} \sum_{x\in B}\kappa_x$. $|\kappa|$

Theorem 2.9 (Laurent Combinatorial Formula for $\mathcal{T}r_{w_1,\ldots,w_\ell}(n)$)**.** *Let $w_1,\ldots,w_\ell \in \mathbf{F}_r$ be a balanced set of words. If $n \ge L_x$ for every $x \in B$, then*

$$\mathcal{T}r_{w_1,\ldots,w_\ell}(n) = \sum_{\sigma\in\overline{\mathrm{MATCH}}^*(w_1,\ldots,w_\ell)} (-1)^{|\kappa(\sigma)|}\, n^{\chi(\sigma)}. \tag{2.10}$$

Proof. This proof relies on grouping the summands in (2.10) according to the "extreme" bijections $\{\sigma_{x,0},\sigma_{x,\kappa_x}\}_{x\in B}$ and show that the total contribution of the summands with extreme bijections $\tau = \{\tau_{x,0},\tau_{x,1}\}_x \in \mathrm{MATCH}^{\kappa\equiv 1}(w_1,\ldots,w_\ell)$ is equal to $\left(\prod_{x\in B}\mathrm{Wg}_{L_x}\left(\tau_{x,0}^{-1}\circ\tau_{x,1}\right)\right)\cdot n^{\#\{o-\text{discs in }\Sigma_\tau\}}$. This is enough by Theorem 2.8.

[9] Novaes [Nov17] has recently obtained a combinatorial formula for the Weingarten function in terms of maps on surfaces; our approach here is different and incorporates that we are integrating over independent unitary matrices, which naturally leads to considerations about infinite groups.

In (2.3) above, substitute $\theta_i = \rho_1 \cdots \rho_i$ to obtain

$$\mathrm{Wg}_L(\pi) = \frac{1}{n^L} \sum_{k\in\mathbb{Z}_{\geq 0}} \sum_{\mathrm{id}=\theta_0\neq\theta_1\neq\ldots\neq\theta_{k-1}\neq\theta_k=\pi} \frac{(-1)^k}{n^{\|\theta_0^{-1}\theta_1\|+\|\theta_1^{-1}\theta_2\|+\ldots+\|\theta_{k-1}^{-1}\theta_k\|}}.$$

Substituting $L = L_x$ and $\pi = \tau_{x,0}^{-1} \cdot \tau_{x,1}$, we get

$$\mathrm{Wg}_{L_x}\left(\tau_{x,0}^{-1} \cdot \tau_{x,1}\right) = \frac{1}{n^{L_x}} \sum_{k\in\mathbb{Z}_{\geq 0}} \sum_{\substack{\theta_0,\ldots,\theta_k\in\mathrm{Sym}\left(x^{+1}-\text{letters of } w_1,\ldots,w_\ell\right) \\ \text{s.t. } \mathrm{id}=\theta_0\neq\theta_1\neq\ldots\neq\theta_{k-1}\neq\theta_k=\tau_{x,0}^{-1}\tau_{x,1}}} \frac{(-1)^k}{n^{\|\theta_0^{-1}\theta_1\|+\|\theta_1^{-1}\theta_2\|+\ldots+\|\theta_{k-1}^{-1}\theta_k\|}}.$$

Multiplying all permutations from the left by $\tau_{x,0}$ and substituting $\sigma_i = \tau_{x,0}\theta_i$, one obtains:

$$\mathrm{Wg}_{L_x}\left(\tau_{x,0}^{-1} \cdot \tau_{x,1}\right) = \frac{1}{n^{L_x}} \sum_{k\in\mathbb{Z}_{\geq 0}} \sum_{\substack{\sigma_0,\ldots,\sigma_k\in\mathrm{MATCH}_x(w_1,\ldots,w_\ell) \\ \text{s.t. } \tau_{x,0}=\sigma_0\neq\sigma_1\neq\ldots\neq\sigma_{k-1}\neq\sigma_k=\tau_{x,1}}} \frac{(-1)^k}{n^{\|\sigma_0^{-1}\sigma_1\|+\|\sigma_1^{-1}\sigma_2\|+\ldots+\|\sigma_{k-1}^{-1}\sigma_k\|}}. \tag{2.11}$$

Note that, by construction, the number of o-discs in Σ_σ depends solely on the extreme bijections in σ, namely on $\{\sigma_{x,0}, \sigma_{x,\kappa_x}\}_{x\in B}$. Thus, together with (2.6) we obtain

$$\begin{aligned}
\mathcal{T}r_{w_1,\ldots,w_\ell}(n) &= \sum_{\tau\in\mathrm{MATCH}^{\kappa\equiv 1}(w_1,\ldots,w_\ell)} \left(\prod_{x\in B} \mathrm{Wg}_{L_x}\left(\tau_{x,0}^{-1} \circ \tau_{x,1}\right)\right) \cdot n^{\#\{o-\text{discs in } \Sigma_\tau\}} \\
&= \sum_{\tau\in\mathrm{MATCH}^{\kappa\equiv 1}} \frac{n^{\#\{o-\text{discs in } \Sigma_\tau\}}}{n^{\sum_x L_x}} \sum_{\kappa\in\left(\mathbb{Z}_{\geq 0}\right)^B} \sum_{\substack{\sigma\in\overline{\mathrm{MATCH}}^\kappa \ s.t. \\ \sigma_{x,0}=\tau_{x,0} \ \& \ \sigma_{x,\kappa_x}=\tau_{x,1}}} \frac{(-1)^{|\kappa|}}{n^{\sum_x\sum_{j=0}^{\kappa_x-1}\|\sigma_{x,j}^{-1}\cdot\sigma_{x,j+1}\|}} \\
&= \sum_{\sigma\in\overline{\mathrm{MATCH}}^*(w_1,\ldots,w_\ell)} (-1)^{|\kappa(\sigma)|}\, n^{\#\{o-\text{discs in } \Sigma_\sigma\}-\sum_x\left(L_x+\sum_{j=0}^{\kappa_x(\sigma)-1}\|\sigma_{x,j}^{-1}\cdot\sigma_{x,j+1}\|\right)}.
\end{aligned}$$

It is thus enough to explain why

$$\chi(\sigma) = \#\{o-\text{discs in } \Sigma_\sigma\} - \sum_x \left(L_x + \sum_{j=0}^{\kappa_x(\sigma)-1} \left\|\sigma_{x,j}^{-1} \cdot \sigma_{x,j+1}\right\|\right).$$

- The number of vertices in Σ_σ is $\sum_{x\in B}(2\kappa_x(\sigma)+4)L_x$ (consisting of $\kappa_x(\sigma)+1$ z-points and a single o-point for each of the $2L_x$ $x^{\pm 1}$-letters in $w_1, \ldots, w_\ell$).
- The number of 1-cells in Σ_σ is $\Sigma_{x\in B}(3\kappa_x(\sigma)+5)L_x$ (there are $\sum_{x\in B}(2\kappa_x(\sigma)+4)L_x$ 1-cells along the boundary components, and additional $\sum_{x\in B}(\kappa_x(\sigma)+1)L_x$ bijection-edges).
- Finally, it is easy to see from Definition 2.7 that the number of cycles in the permutation $\sigma_{x,j}^{-1} \cdot \sigma_{x,j+1}$ is identical to the number of (x,j)-discs in Σ_σ, so that

$$\left\|\sigma_{x,j}^{-1} \cdot \sigma_{x,j+1}\right\| = L_x - \#\{(x,j)-\text{discs in } \Sigma_\sigma\}.$$

Therefore,

$$\begin{aligned}
\chi(\sigma) &= \left[\sum_{x\in B}(2\kappa_x(\sigma)+4)L_x\right]-\left[\sum_{x\in B}(3\kappa_x(\sigma)+5)L_x\right]+\\
&+\left[\sum_{x\in B}\sum_{j=0}^{\kappa_x(\sigma)-1}\#\{(x,j)-\text{discs in }\Sigma_\sigma\}\right]+\#\{o-\text{discs in }\Sigma_\sigma\}\\
&= \#\{o-\text{discs in }\Sigma_\sigma\}+\sum_{x\in B}\left[(-\kappa_x(\sigma)-1)L_x+\sum_{j=0}^{\kappa_x(\sigma)-1}\left(L_x-\left\|\sigma_{x,j}^{-1}\cdot\sigma_{x,j+1}\right\|\right)\right]\\
&= \#\{o-\text{discs in }\Sigma_{\tilde\sigma}\}-\sum_{x\in B}\left[L_x+\sum_{j=0}^{\kappa_x(\sigma)-1}\left\|\sigma_{x,j}^{-1}\cdot\sigma_{x,j+1}\right\|\right]. \qquad (2.12)
\end{aligned}$$

□

It is implicit in Theorem 2.9 and its proof that there are only finitely many sets of sequences of bijections σ with contribution of a given order. Namely, for every integer χ_0 there are finitely many σ in the summation (2.10) with $\chi(\sigma)=\chi_0$. This is true because the same property holds for the asymptotic expansion of the Weingarten function in (2.3). However, for completeness, we give a direct proof for this fact:

Claim 2.10. For every $\chi_0\in\mathbb{Z}$ there are finitely many sets σ in the sum (2.10) with $\chi(\sigma)=\chi_0$.

Proof. The number of o-discs in Σ_σ is bounded by the number of o-points in $S^1(w_1)\cup\ldots\cup S^1(w_\ell)$. All sets σ in the sum (2.10) satisfy $\sigma_{x,j}\neq\sigma_{x,j+1}$ for all $x\in B$ and $0\le j\le\kappa_x-1$, and so $\left\|\sigma_{x,j}^{-1}\cdot\sigma_{x,j+1}\right\|\ge 1$. Thus, from (2.12) we obtain that if $\chi(\sigma)=\chi_0$ then

$$\chi_0=\chi(\sigma)\le\#\left\{o-\text{points in }S^1(w_1)\cup\ldots\cup S^1(w_\ell)\right\}-\sum_{x\in B}\left[L_x+\kappa_x(\sigma)\right].$$

Hence

$$\sum_{x\in B}\kappa_x(\sigma)\le\#\left\{o-\text{points in }S^1(w_1)\cup\ldots\cup S^1(w_\ell)\right\}-\chi_0-\sum_{x\in B}L_x.$$

Since the right hand side is independent of σ, the proof is completed. □

The duality between the two types of formulas in Theorems 2.8 and 2.9 will be manifested also in the next section. Our main object of study will be the complex $\mathcal{T}(\Sigma,f)$ of transverse maps which, similarly to the sets in the infinite formula (2.10), consists of sequences of arcs and curves of arbitrary lengths, but with the single constraint that two consecutive objects in every sequence must be different from each other (strict transverse maps – see Definition 3.3). However, for one of the main results about $\mathcal{T}(\Sigma,f)$, namely, its being contractible, we return to the model of sequences of length two without the constraint of two consecutive objects being different – see Definition 3.15.

2.5 Maps from the surfaces to the bouquet

In Definition 2.7 and in Theorems 2.8 and 2.9 we introduced surfaces associated with $w_1,\ldots,w_\ell$ and tuples of matchings. The following definition introduces a natural map from these surfaces to the bouquet so that each surface and its associated map turns into an admissible pair for $w_1,\ldots,w_\ell$.

Definition 2.11. Let $w_1,\ldots,w_\ell\in\mathbf{F}_r$, let $\sigma\in\mathrm{MATCH}^\kappa(w_1,\ldots,w_\ell)$ for some $\kappa\in(\mathbb{Z}_{\ge0})^B$ and let Σ_σ be the surface constructed in Definition 2.7. Define $f_\sigma\colon\Sigma_\sigma\to\bigvee^r S^1$ as follows: f_σ

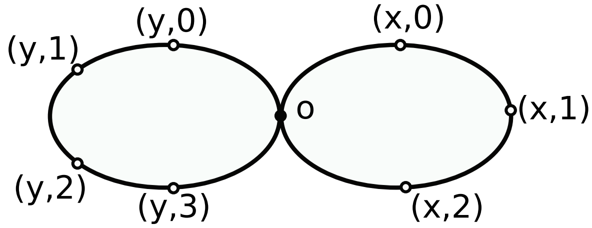

Figure 2.3: The wedge $\bigvee^r S^1$ with transversion points. Here $r = 2$, $B = \{x, y\}$, $\kappa_x = 2$ and $\kappa_y = 3$.

- For every $x \in B$, mark $\kappa_x + 1$ distinct points on the circle of the bouquet $\bigvee^r S^1$ corresponding to x, away from the wedge point o, and label them $(x, 0), \ldots, (x, \kappa_x)$ in the order of the orientation of the circle. See Figure 2.3.
- The preimage through f_σ^{-1} of $(x, j) \in \bigvee^r S^1$ is exactly the bijection-edges corresponding to $\sigma_{x,j}$, which contain the (x, j)-points of Σ_σ as their endpoints.
- The o-points in Σ_σ are mapped to $o \in \bigvee^r S^1$.
- On $S^1(w_i)$, on each of the $|w|$ intervals, if the interval I corresponds to the letter x^ε, $\varepsilon \in \{\pm 1\}$, $f_\sigma\Big|_I$ traces the x-circle in $\bigvee^r S^1$ monotonically, with orientation prescribed[10] by ε.
- Finally, for every open disc D in the CW-complex Σ_σ, the image of f_σ along ∂D is nullhomotopic, so there is a unique way to extend it to D, up to homotopy, and we extend it so that the image of f on the interior of D avoids the marked points $\{(x, j)\}_{x \in B, j \in [\kappa_x]} \subset \bigvee^r S^1$.

It is evident that $(\Sigma_\sigma, f_\sigma)$ is admissible for $w_1, \ldots, w_\ell$ (with the appropriate o-point in $S^1(w_i)$ labeled also as v_i, for every $i = 1, \ldots, \ell$). In particular,

Corollary 2.12. *If $w_1 \cdots w_\ell \in [\mathbf{F}_r, \mathbf{F}_r]$ then $\mathsf{Surfaces}(w_1, \ldots, w_\ell) \neq \emptyset$.*

Another important observation is that every incompressible pair $[(\Sigma, f)] \in \mathsf{Surfaces}(w_1, \ldots, w_\ell)$ has a representative in the form of $(\Sigma_\sigma, f_\sigma)$ with only one matching per letter:

Proposition 2.13. *Denote by $\mathrm{MATCH}^{\kappa \equiv 0}(w_1, \ldots, w_\ell)$ the set of matchings corresponding to $\kappa_x = 0$ for every $x \in B$. Every incompressible pair (Σ, f) which is admissible for $w_1, \ldots, w_\ell$ is equivalent to $(\Sigma_\sigma, f_\sigma)$ for some $\sigma \in \mathrm{MATCH}^{\kappa \equiv 0}(w_1, \ldots, w_\ell)$.*

Proof. The argument here imitates the one in [Cul81, Theorem 1.4]. Assume (Σ, f) is incompressible. Mark a point $(x, 0)$ on the middle of the circle corresponding to x in $\bigvee^r S^1$, and perturb f (relative to the points $v_1, \ldots, v_\ell \in \Sigma$) so that $f(\partial_i \Sigma)$ is monotone, namely, never backtracking, for $i = 1, \ldots, \ell$, and so that it becomes transverse to $\{(x, 0)\}_{x \in B} \subset \bigvee^r S^1$ (see Definition 3.1 below). As f is transverse to $(x, 0) \in \bigvee^r S^1$, the preimage of $(x, 0)$ consists of a collection of disjoint arcs and curves (in this paper we use the notion "curve" as synonym for "simple closed curve"). Because $f(\partial_i \Sigma)$ is monotone, there are exactly L_x arcs, which determine an element $\sigma \in \mathrm{MATCH}^{\kappa \equiv 0}(w_1, \ldots, w_\ell)$. There are no curves in $f^{-1}((x, 0))$ because such curves would be null-curves, which is impossible with f being incompressible. Finally, f being incompressible also guarantees that the collection of arcs $\coprod_{x \in B} f^{-1}((x, 0))$ cuts Σ into discs. Thus $(\Sigma, f) \sim (\Sigma_\sigma, f_\sigma)$. $\square$

Since the set $\mathrm{MATCH}^{\kappa \equiv 0}(w_1, \ldots, w_\ell)$ is finite, we obtain:

Corollary 2.14. *There are finitely many classes of incompressible (Σ, f) in $\mathsf{Surfaces}(w_1, \ldots, w_\ell)$.*

[10] We mention this specifically because when $\kappa_x = 0$ this does not follow from the previous bullet points.

In Section 5.2 we address the issue of how one can distinguish the different incompressible classes in $\mathsf{Surfaces}\left(w_1,\ldots,w_\ell\right)$.

At this point we can also derive the asymptotic bounds we have for $\mathcal{T}r_{w_1,\ldots,w_\ell}(n)$:

Proof of Corollary 1.8. Recall that we need to prove that $\mathcal{T}r_{w_1,\ldots,w_\ell}(n) = O\left(n^{\chi_{\max}(w_1,\ldots,w_\ell)}\right)$ where $\chi_{\max}\left(w_1,\ldots,w_\ell\right)$ is the maximal Euler characteristic of a surface in $\mathsf{Surfaces}\left(w_1,\ldots,w_\ell\right)$. Theorem 2.8 says that $\mathcal{T}r_{w_1,\ldots,w_\ell}(n)$ is equal to a sum over $\sigma \in \mathrm{MATCH}^{\kappa\equiv 1}\left(w_1,\ldots,w_\ell\right)$, and the contribution of each σ is $c\cdot n^{\chi(\sigma)} + O\left(n^{\chi(\sigma)-2}\right)$ where $c \in \mathbb{Z}\setminus\{0\}$. As $\left(\Sigma_\sigma, f_\sigma\right) \in \mathsf{Surfaces}\left(w_1,\ldots,w_\ell\right)$, then by definition $\chi(\sigma) \le \chi_{\max}\left(w_1,\ldots,w_\ell\right)$. $\square$

Remark 2.15. In fact, we get even more: every $[(\Sigma,f)] \in \mathsf{Surfaces}\left(w_1,\ldots,w_\ell\right)$ attaining $\chi_{\max}\left(w_1,\ldots,w_\ell\right)$ is incompressible, and therefore, by Proposition 2.13, equivalent to $\left(\Sigma_\sigma, f_\sigma\right)$ for some
$\sigma \in \mathrm{MATCH}^{\kappa\equiv 0}\left(w_1,\ldots,w_\ell\right)$. This σ takes part in the expression for $\mathcal{T}r_{w_1,\ldots,w_\ell}(n)$ in Theorem 2.9 and thus one can expect that $\mathcal{T}r_{w_1,\ldots,w_\ell}(n) = \Theta\left(n^{\chi_{\max}(w_1,\ldots,w_\ell)}\right)$. As some of the examples from Table 1 indicate, the coefficient of $n^{\chi_{\max}(w_1,\ldots,w_\ell)}$ may vanish, but this only happens if the different non-zero contributions cancel out. (One can get to the same conclusion from the finite formula for $\mathcal{T}r_{w_1,\ldots,w_\ell}(n)$ in Theorem 2.8, by duplicating every matching in $\sigma \in \mathrm{MATCH}^{\kappa\equiv 0}$ to obtain $\sigma' \in \mathrm{MATCH}^{\kappa\equiv 1}\left(w_1,\ldots,w_\ell\right)$, which then satisfies $\left(\Sigma_\sigma, f_\sigma\right) \sim \left(\Sigma_{\sigma'}, f_{\sigma'}\right)$.)

Reduction of the main theorem

Recall that Theorem 2.9 expresses $\mathcal{T}r_{w_1,\ldots,w_\ell}(n)$ as a sum over the (generally infinite) set $\overline{\mathrm{MATCH}}^*\left(w_1,\ldots,w_\ell\right)$. To prove our main result, Theorem 1.7, we group together all $\sigma \in \overline{\mathrm{MATCH}}^*\left(w_1,\ldots,w_\ell\right)$ for which $\left(\Sigma_\sigma, f_\sigma\right)$ belong to the same equivalence class, and show the total contribution of these values of σ to (2.10) is exactly the one specified in Theorem 1.7. Accordingly, for $[(\Sigma,f)] \in \mathsf{Surfaces}$ we let $\overline{\mathrm{MATCH}}^*\left(w_1,\ldots,w_\ell;\Sigma,f\right)$ be the σ's yielding elements in the class of (Σ,f). So, recalling the notation "$\sim$" from Definition 1.3, $\overline{\mathrm{MATCH}}^*(\ldots;\Sigma,f)$

$$\overline{\mathrm{MATCH}}^*\left(w_1,\ldots,w_\ell;\Sigma,f\right) \stackrel{\mathrm{def}}{=} \left\{\sigma \in \overline{\mathrm{MATCH}}^*\left(w_1,\ldots,w_\ell\right) \,\middle|\, \left(\Sigma_\sigma, f_\sigma\right) \sim (\Sigma,f)\right\}.$$

From Claim 2.10 it follows that $\overline{\mathrm{MATCH}}^*\left(w_1,\ldots,w_\ell;\Sigma,f\right)$ is finite for every $(\Sigma,f) \in \mathsf{Surfaces}\left(w_1,\ldots,w_\ell\right)$. Using Theorem 2.9, Theorems 1.4 and 1.7 now reduce to:

Theorem 2.16. *Let* $[(\Sigma,f)] \in \mathsf{Surfaces}\left(w_1,\ldots,w_\ell\right)$. *Then* $\chi^{(2)}\left(\mathrm{MCG}(f)\right)$ *is well defined and given by*

$$\chi^{(2)}\left(\mathrm{MCG}(f)\right) = \sum_{\sigma \in \overline{\mathrm{MATCH}}^*\left(w_1,\ldots,w_\ell;\Sigma,f\right)} (-1)^{|\kappa(\sigma)|}. \tag{2.13}$$

In particular, if $[(\Sigma,f)] \in \mathsf{Surfaces}\left(w_1,\ldots,w_\ell\right)$ cannot be realized by any $\sigma \in \overline{\mathrm{MATCH}}^*\left(w_1,\ldots,w_\ell\right)$, then $\chi^{(2)}\left(\mathrm{MCG}(f)\right) = 0$. Since there are only finitely many σ with Σ_σ of a given Euler characteristic (Claim 2.10), it follows from Theorem 2.16 that, indeed, for any $\chi_0 \in \mathbb{Z}$, there are only finitely many classes $[(\Sigma,f)] \in \mathsf{Surfaces}\left(w_1,\ldots,w_\ell\right)$ with $\chi(\Sigma) = \chi_0$ and for which $\chi^{(2)}\left(\mathrm{MCG}(f)\right) \neq 0$.

In the next two sections we describe the constructions and results that lead to the proof of Theorem 2.16 and of Theorem 1.6 which strengthen the result in the case (Σ,f) is incompressible. We hint that the special property of the incompressible case is that the set $\overline{\mathrm{MATCH}}^*\left(w_1,\ldots,w_\ell;\Sigma,f\right)$ gives rise to a natural complex with one cell for every element $\sigma \in \overline{\mathrm{MATCH}}^*\left(w_1,\ldots,w_\ell;\Sigma,f\right)$, so that the Euler characteristic of this complex is exactly the right hand side of (2.13). See Section 4.4 for details.

3 A complex of transverse maps

The key ingredient in the proof of our main results is a complex of transverse maps which we associate with a given pair $[(\Sigma, f)] \in \mathsf{Surfaces}(w_1, \ldots, w_\ell)$. In the current section we define it, study important properties and prove it is contractible. In the next section we study the action of $\mathrm{MCG}(f)$ on this complex and prove our main results.

3.1 Transverse maps

Recall that in this paper the term "curve" is short for a simple closed curve.

Definition 3.1. Let Σ be orientable. A map $f\colon \Sigma \to \bigvee^r S^1$ is said to be *transverse* to a point $p \in \bigvee^r S^1 \setminus \{o\}$ if the preimage of p is a disjoint union of arcs and curves, and if in a small tubular neighborhood U of every curve or arc γ in the preimage, the two connected components of $U \setminus \gamma$ are mapped to two different "sides" of p. transverse

For example, the map f_σ from Definition 2.11 is transverse to the points $\{(x,j)\}_{x\in B, j\in[\kappa_x]}$ in $\bigvee^r S^1$. In this case, the preimage of each of these points contains no curves but rather only arcs. We shall consider here different realizations of the homotopy class $[f]$ of the same map $f\colon \Sigma \to \bigvee^r S^1$, which are transverse to different collections of points in $\bigvee^r S^1$.

More formally, let Σ be a surface and f a map $f\colon \Sigma \to \bigvee^r S^1$ so that $[(\Sigma, f)] \in \mathsf{Surfaces}(w_1, \ldots, w_\ell)$. By definition, Σ has ℓ marked points: one point, labeled v_i, in every boundary component $\partial_i\Sigma$, for $i = 1, \ldots, \ell$, and with $f(v_i) = o$. Note that $w_1, \ldots, w_\ell$ are prescribed from Σ and f by $w_i = f_*\left(\overrightarrow{\partial_i\Sigma}, v_i\right) \in \pi_1\left(\bigvee^r S^1, o\right)$. For every $i = 1, \ldots, \ell$, we mark additional $|w_i| - 1$ points on $\partial_i\Sigma$ inside $f^{-1}(o)$, so that f maps the intervals of $\partial_i\Sigma$ cut by these points to the letters of w_i. We let $V_o \subset \Sigma$ denote the set of all marked points in Σ: a total of $\sum_{i=1}^{\ell} |w_i|$ marked points all at the boundary. V_o

Definition 3.2. Let $\kappa = \{\kappa_x\}_{x\in B} \in (\mathbb{Z}_{\geq 0})^B$ be a set of non-negative integers. On the circle corresponding to x in $\bigvee^r S^1$ mark $\kappa_x + 1$ disjoint points, $(x,0), \ldots, (x,\kappa_x)$, arranged as in Definition 2.11 and Figure 2.3. Let $[(\Sigma, f)] \in \mathsf{Surfaces}(w_1, \ldots, w_\ell)$ and $V_o \subset f^{-1}(o) \subseteq \Sigma$ be as above. A map $g\colon \Sigma \to \bigvee^r S^1$ is *a transverse map realizing* (Σ, f) *with parameters* κ, if it is homotopic to f relative to V_o and transverse to the points $\{(x,j)\}_{x\in B, j\in[\kappa_x]} \subset \bigvee^r S^1$. Note that, in particular, $g(V_o) = \{o\}$.

An arc (curve, respectively) in the preimage of (x,j) is called an (x,j)-arc ((x,j)-curve). Let U_o be the connected component of o in $\bigvee^r S^1 \setminus \{(x,j)\}_{x\in B, j\in[\kappa_x]}$. A connected component of $g^{-1}(U_o)$ is called an *o-zone*. For $0 \leq j \leq \kappa_x - 1$, let $I_{x,j} \subset \bigvee^r S^1$ be the interval on the x-circle cut out by (x,j) and $(x,j+1)$. We call a connected component of $g^{-1}(I_{x,j})$ an (x,j)-*zone*, or, if x and j are not relevant, also a z-zone. If all zones defined by g are topological discs, we say that g *fills* Σ, or that g is *filling*. (x,j)-arc/curve o-zone (x,j)-zone z-zone fills

The isotopy[11] class of the transverse map g, denoted $[g]$, contains all transverse maps with the same parameters κ which are homotopic to g relative to V_o via a homotopy of transverse maps with the same parameters. We stress that the marked points in $\bigvee^r S^1$ are allowed to move inside $\bigvee^r S^1 \setminus \{o\}$ along the homotopy as long as they remain disjoint. $[g]$

Note that every (x,j)-arc/curve has a direction from one side of the arc/curve to the other, induced by the orientation of the circle in $\bigvee^r S^1$. Since we do not care about the location of the (x,j)-points in $\bigvee^r S^1$, a transverse map g for (Σ, f) can be identified with the collection of disjoint "directed" and colored arcs and curves. The isotopy class of g can then be though of as the isotopy

[11] We call $[g]$ the isotopy class of g, rather than the homotopy class, because if one thinks of g as a collection of disjoint colored arcs and curves embedded in Σ, then $[g]$ is indeed the isotopy class of this collection relative to V_o.

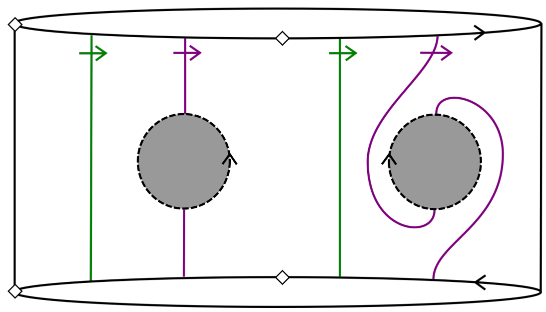

Figure 3.1: A collection of arcs corresponding to a transverse map realizing (Σ, f), where Σ is a genus 1 surface with 2 boundary components, drawn as an annulus with two discs cut out, with boundaries of those discs identified. Here $r = 1$ and there is one generator x. Green arcs are $(x, 0)$-arcs and purple are $(x, 1)$-arcs. The marked points V_o are diamonds. There are no curves in this system and it is filling. The words at the boundary are x^2, x^{-2}.

class of this collection of arcs and curves relative to V_o. We illustrate such a collection in Figure 3.1.

Also note, by the definition above, that the boundary of an (x, j)-zone of g consists of pieces of $\partial\Sigma$, of (x, j)-arcs/curves directed inward and of $(x, j+1)$-arcs/curves directed outward. In contrast, the boundary of an o-zone of g consists of pieces of $\partial\Sigma$, of $(x, 0)$-arcs/curves directed outward and of (x, κ_x)-arcs/curves directed inward, for various $x \in B$. Finally, every point in V_o belongs to some o-zone of g.

Generally, we want to forbid certain trivial or redundant features of transverse maps, as we elaborate in the following definition:

Definition 3.3. A transverse map for (Σ, f) is called *loose* if it satisfies loose transverse map

- **Restriction 1**. There are no o-zones nor z-zones that contain no marked point from V_o and whose boundary arcs and curves have the same color (x, j) and are all oriented pointing inwards or all oriented outwards. Note this rules out the possibility of a zone that is a disc bounded by a closed curve.

- **Restriction 2**. No segment of the boundary of Σ that contains no marked point can be bounded by the end points of two arcs that are equally-labeled and both directed inwards or both outwards. Note that this is the boundary analog of **Restriction 1**.

A transverse map for (Σ, f) is called *strict* if it satisfies, in addition, strict transverse map

- **Restriction 3**. For every $x \in B$ and $0 \le j \le \kappa_x - 1$, the collection of (x, j)-arcs and curves is not isotopic to the collection of $(x, j+1)$-arcs and curves. In other words, there must be at least one (x, j)-zone which is neither a rectangle nor an annulus[12].

Remark 3.4.
- Note that if g fills Σ then there are no curves involved in g, but only arcs. This is the case, for example, when $g = f_\sigma$ as in Definition 2.11.

- Any transverse map for (Σ, f) satisfying **Restriction 2** admits exactly $\kappa_x + 1$ arcs touching every interval in $\partial\Sigma$ corresponding to the letter x, one arc for every $j \in [\kappa_x]$. Consequently, it admits exactly L_x arcs labeled (x, j) for every $x \in B$ and $j \in [\kappa_x]$. In addition, if O is an o-zone of such a map then every connected component of $\overline{O} \cap \partial\Sigma$ contains exactly one marked point from V_o.

[12]Here, a rectangle is a disc bounded by two arcs and two pieces of $\partial\Sigma$, and an annulus is bounded by two curves. Restriction 3 should resonate the constraint on the set of matchings $\overline{\text{MATCH}}^{\kappa}(w_1, \ldots, w_\ell)$ from Section 2.4. In particular, if $g(\Sigma)$ does not contain the circle in $\bigvee^r S^1$ associated with x, then necessarily $\kappa_x = 0$.

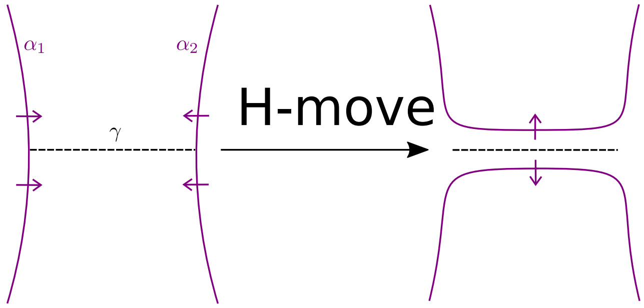

Figure 3.2: The dashed line is the arc-segment γ along which the H-move is performed. The two purple segments on the left are parts of α_1 and α_2.

- If **Restriction 2** holds, then **Restriction 1** can only fail at zones bounded by curves.

The following local surgery of a transverse map will be very useful in the sequel:

Definition 3.5. Let g be a transverse map realizing (Σ, f) with parameters κ. Let α_1 and α_2 be each an (x, j)-arc or an (x, j)-curve in the collection corresponding to g, where α_1 and α_2 have the same color and are not necessarily distinct. Assume further there is an embedded arc-segment γ inside the interior of Σ, with one endpoint in α_1 and the other in α_2 such that the interior of γ is disjoint from the arc-curve collection of g, and such that both α_1 and α_2 are directed towards γ or both directed away from γ. We say that the transverse map g' realizing (Σ, f) with the same parameters is obtained from g by an *H-Move* along γ if: H-Move

- One takes a small collar neighborhood of γ to obtain a rectangle whose short sides are contained in α_1 and in α_2.
- One deletes the short sides of the rectangle from the arc-curve collection of g and replaces them with the long sides to obtain a new collection which defines g'. See Figure 3.2.

It is clear that g' is homotopic to g as a map but not isotopic as a transverse map.

3.2 The poset of strict transverse maps

The complex of transverse maps will be defined as the geometric realization of the *poset* of transverse maps:

Definition 3.6. Let Σ and $f\colon \Sigma \to \bigvee^r S^1$ satisfy $[(\Sigma, f)] \in \mathsf{Surfaces}\,(w_1, \ldots, w_\ell)$. Let $V_o \subset f^{-1}(o) \cap \partial\Sigma$ be defined as above, so $V_o \cap \partial_i\Sigma$ cuts $\partial_i\Sigma$ into $|w_i|$ intervals, each of which is mapped to some $x^{\pm 1}$ with $x \in B$ by f_*. The *poset of transverse maps realizing* (Σ, f), denoted $\mathcal{T} = \mathcal{T}(\Sigma, f)$ or $(\mathcal{T}, \preceq)$, consists of the set of isotopy classes relative to V_o of ***strict*** transverse maps realizing (Σ, f). The order is defined by "forgetting points of transversion". Namely, whenever g_1 is a transverse map realizing (Σ, f) with parameters $\{\kappa_x\}_{x\in B}$ and g_2 is identical to g_1 except we forget a proper (possibly empty) subset of the transversion points $\{(x, j)\}_{j\in[\kappa_x]} \subset \bigvee^r S^1$ for every $x \in B$, then the isotopy classes $[g_1]$ and $[g_2]$ satisfy $[g_2] \preceq [g_1]$ in the poset $\mathcal{T}$. $\mathcal{T}{=}\mathcal{T}(\Sigma,f)$

Of course, the transversion points that remain in g_2 may need relabeling. Note that we use here *strict* transverse maps: maps satisfying the three restrictions from Definition 3.3. The role of *loose* transverse maps will be clarified in the sequel of this section. Another important observation is that if the transverse map g is strict, then so are the maps obtained from g by forgetting transversion points:

Lemma 3.7. *If g_1 is a strict transverse map realizing (Σ, f) and g_2 is obtained from g_1 by forgetting transversion points, then g_2 is also strict. In other words, if $[g_1] \in \mathcal{T}$ then $[g_2] \in \mathcal{T}$.*

Proof. It is enough to prove the statement of the lemma in the special case where g_2 is obtained by forgetting a single point, say the point (x, j), for some $x \in B$ with $\kappa_x \geq 1$ and $j \in [\kappa_x]$. It is obvious that g_2 satisfies **Restriction 3**. Let $I \subset \partial_i \Sigma \setminus V_o$ be an interval cut out by two adjacent marked points. As g_1 satisfies **Restriction 2**, all arcs touching I are directed in the same orientation w.r.t. I and this property remains true for g_2, hence g_2 satisfies **Restriction 2**.

As for **Restriction 1**, assume first that $j = \kappa_x$, so every o-zone and $(x, \kappa_x - 1)$-zone of g_1 which are neighbors belong to the same o-zone of g_2. Every z-zone of g_2 is also a z-zone of g_1 with the same boundary, so **Restriction 1** is not violated there. Let $O \subset \Sigma$ be an o-zone of g_2 that violates **Restriction 1**. Then O is not an o-zone of g_1, and has to be a union of o-zones and $(x, \kappa_x - 1)$-zones of g_1, at least one of each type. In addition, O contains no marked points and has only incoming $(x, \kappa_x - 1)$-arcs/curves at its boundary: this is because g_1 satisfies **Restriction 1**, every $(x, \kappa_x - 1)$-zone has some incoming $(x, \kappa_x - 1)$-arc/curve at its boundary, and so O also has some incoming $(x, \kappa_x - 1)$-arc/curve at its boundary. But then, every o-zone of g_1 contained in O has no marked points and only incoming (x, κ_x)-arcs/curves at its boundary, a contradiction.

The proof is analogous if $j = 0$ and is similar but even simpler if $1 \leq j \leq \kappa_x - 1$. $\square$

Another important observation is that $\mathcal{T} = \mathcal{T}(\Sigma, f)$ is not empty.

Lemma 3.8. *Let Σ, f be as in Definition 3.6, then the poset $\mathcal{T} = \mathcal{T}(\Sigma, f)$ is not empty.*

Proof. For every $x \in B$, mark a single point $(x, 0)$ in $\bigvee^r S^1$ on the circle corresponding to x. Perturb f to obtain g that is transverse to the points $\{(x, 0)\}_{x \in B}$ (without changing the image at V_o). The resulting map is a transverse map realizing (Σ, f) with parameters $\kappa_x = 0$ for all x. **Restriction 3** is automatically satisfied when $\kappa_x = 0$ for all x.

If g violates **Restriction 2**, then there is a segment I of the boundary cut out by two endpoints of $(x, 0)$-arcs for some $x \in B$, both directed, say, inwards, and without any marked point. Let γ be an arc parallel to I slightly away from $\partial\Sigma$ with endpoint at the two arcs cutting I. Perform an H-move along γ, and delete the resulting $(x, 0)$-arc parallel to I and γ. In that manner one can get rid of all violations of **Restriction 2**.

So assume now that g does not violate **Restrictions 2 and 3**. Any violation of **Restriction 1** is at zones bounded by curves. But any such zone can be simply deleted by removing all its bounding curves. To see that this procedure does not change the homotopy type of the function, note that it can be achieved by a series of H-moves: first perform H-moves along arcs connecting a curve and itself to decrease the genus of the zone to 0. Then, use H-moves between different bounding curves to eventually reduce the number of bounding curves to one. The resulting zone is a disk bounded by a curve which can easily be homotoped away. We can repeat this process until no violations of **Restriction 3** remain. The resulting map is a strict transverse map realizing (Σ, f). $\square$

3.3 The complex of transverse maps

The complex of transverse maps is defined as a "polysimplicial complex", meaning that its cells are products of simplices*, or polysimplices*, as in $\Delta_{k_1} \times \Delta_{k_2} \times \ldots \times \Delta_{k_r}$, where Δ_k is the standard simplex of dimension k. Note that the polysimplex $\Delta_{k_1} \times \ldots \times \Delta_{k_r}$ has dimension $k_1 + \ldots + k_r$. polysimplex

Definition 3.9. *The complex of transverse maps realizing (Σ, f), denoted* $|\mathcal{T}|_{\text{poly}} = |\mathcal{T}(\Sigma, f)|_{\text{poly}}$, is a polysimplicial complex with a polysimplex $\text{polysim}([g]) \overset{\text{def}}{=} \prod_x \Delta_{\kappa_x}$ for every element $[g] \in \mathcal{T}$ with parameters $\{\kappa_x\}_{x \in B}$. The faces of $\text{polysim}([\text{g}])$ are exactly $\{\text{polysim}([g']) \mid [g'] \preceq [g]\}$. Then $|\mathcal{T}|_{\text{poly}}$ is the union of closed cells or disjoint union of open cells: $|\mathcal{T}|_{\text{poly}}$ $\text{polysim}([g])$

$$|\mathcal{T}|_{\text{poly}} \overset{\text{def}}{=} \bigcup_{[g] \in \mathcal{T}} \overline{\text{polysim}}([\text{g}]) = \bigsqcup_{[g] \in \mathcal{T}} \text{polysim}^{\text{o}}([\text{g}])\,.$$

The topology on $|\mathcal{T}|_{\mathrm{poly}}$, as the topology on every (poly-)simplicial complex in this paper, is defined by taking the Euclidean topology on every (poly-)simplex s, and by letting a general set $A \subseteq |\mathcal{T}|_{\mathrm{poly}}$ to be closed if and only if $A \cap s$ is closed in s for every (poly-)simplex s.

We remark that **Restriction 3** plays an important role in this definition: it guarantees that different vertices of the closed polysimplex $\overline{\mathrm{polysim}}\,([g])$ correspond to different (minimal) elements of $\mathcal{T}$, hence the closed polysimplices are embedded in $|\mathcal{T}|_{\mathrm{poly}}$.

There is an equivalent way to construct the complex of transverse map (up to homeomorphism), as an ordinary simplicial complex: the *order complex* $|\mathcal{T}|$ of $\mathcal{T}$. This is a standard simplicial complex, with simplices corresponding to *chains* in $\mathcal{T}$: every chain $[g_0] \prec [g_1] \prec \ldots \prec [g_m]$ corresponds to an m-simplex, with the obvious faces. $|\mathcal{T}|$

Claim 3.10. $|\mathcal{T}|$ is the barycentric subdivision of $|\mathcal{T}|_{\mathrm{poly}}$. In particular, $|\mathcal{T}| \cong |\mathcal{T}|_{\mathrm{poly}}$.

To prove the claim we use the following well-known fact. Here, if $(P, \leq_P)$ and $(Q, \leq_Q)$ are posets, then $|P|$ is the order complex of P, and the direct product $(P \times Q, \leq_{P\times Q})$ is defined by $(p_1, q_1) \leq_{P\times Q} (p_2, q_2)$ if and only if $p_1 \leq_P p_2$ and $q_1 \leq_Q q_2$.

Fact 3.11 (e.g. [Wal88, Theorem 3.2]). *Let P and Q be posets. The function $|P \times Q| \to |P| \times |Q|$ defined by*

$$\sum \lambda_i\,(p_i, q_i) \mapsto \left(\sum \lambda_i p_i, \sum \lambda_i q_i\right)$$

is an homeomorphism.

Proof of Claim 3.10. Let $[g] \in \mathcal{T}$ with parameters $\kappa \in (\mathbb{Z}_{\geq 0})^B$. We show that the barycentric subdivision of $\overline{\mathrm{polysim}}\,([g])$ consists of the simplices corresponding to chains in $\mathcal{T}$ with top element $[g']$ satisfying $[g'] \preceq [g]$. Indeed, this is certainly true in the single-letter case where $r = |B| = 1$ and every polysimplex is merely a simplex. For the general case, let $[\gamma^x(g)]$ denote the isotopy class of the collection of arcs/curves corresponding to the letter x, for $x \in B$. Let $P_x(g)$ denote the poset of all isotopy classes of collections of arcs/curves obtained from $[\gamma^x(g)]$ by forgetting arcs/curves from a proper subset of the colors $[\kappa_x]$. The single-letter case shows that $|P_x(g)| \cong \Delta_{\kappa_x}$. Since the subposet of $\mathcal{T}$ given by $\mathcal{T}_{\preceq[g]} \stackrel{\mathrm{def}}{=} \{[g'] \in \mathcal{T} \mid [g'] \preceq [g]\}$ is exactly $\prod_{x\in B} P_x(g)$, Fact 3.11 yields that the order complex of $\mathcal{T}_{\preceq[g]}$ is homeomorphic to $\overline{\mathrm{polysim}}\,([g])$. □

An important property of $|\mathcal{T}|_{\mathrm{poly}}$ is that it is finite-dimensional. This is an analog of Claim 2.10 and here, again, **Restriction 3** plays an important role:

Lemma 3.12. *The complex $|\mathcal{T}|_{\mathrm{poly}}$ is finite dimensional with*[13] $\dim\left(|\mathcal{T}|_{\mathrm{poly}}\right) \leq \frac{\ell}{2} - \chi(\Sigma)$.

Proof. We need to show that $\sum_x \kappa_x$ is bounded across $[g] \in \mathcal{T}$. It is easy to see that

$$\chi(\Sigma) = \sum_{\Sigma'} \left(\chi(\Sigma') - \frac{1}{2}\#\left\{\text{arcs at } \partial\Sigma'\right\}\right), \tag{3.1}$$

where the sum is over all o-zones and z-zones of g in Σ, and an arc that bounds Σ' from both its sides is counted twice for Σ'. The contribution of Σ' in (3.1) is positive only if Σ' is a topological disc with at most one arc at its boundary. By **Restrictions 1** and **2** this means that Σ' is bounded by one arc and one interval from $\partial\Sigma$ containing a marked point, and that its contribution is $\frac{1}{2}$. Notice that in this case, the marked point must be the special point $v_i \in \partial_i\Sigma$ marking "the beginning" of w_i, and w_i must be not cyclically reduced. Hence the positive contributions on the right hand side of (3.1) sum up to at most $\frac{\ell}{2}$, and come from o-zones only.

[13] Recall Remark 1.14 that we assume $w_i \neq 1$ throughout the proofs. If we do consider the case that some of the words are trivial, then Σ may contain components made of discs, and the bound in Lemma 3.12 needs to be updated.

On the other hand, the only zones contributing zero to (3.1) are discs with two arcs at their boundary, namely, rectangles, or annuli bounded by two curves. Every other z-zone contributes at most -1: this follows from the fact that every boundary component of such a zone is either a curve or contains an even number of arcs. Thus, **Restriction 3** guarantees that for every $x \in B$ and $0 \leq j \leq \kappa_x - 1$, the total contribution of the (x, j)-zones is at most -1. We obtain

$$\chi(\Sigma) \leq \frac{\ell}{2} - \sum_x \kappa_x,$$

so

$$\sum_x \kappa_x \leq \frac{\ell}{2} - \chi(\Sigma).$$

□

Remark 3.13. When $w_1, \ldots, w_\ell$ are cyclically reduced and none equal to 1, the proof gives $\dim\left(|\mathcal{T}|_{\mathrm{poly}}\right) \leq -\chi(\Sigma)$.

The following theorem is the main result of the current section. It is established in Section 3.5.

Theorem 3.14. *The complex of transverse maps $|\mathcal{T}|_{\mathrm{poly}}$ is contractible.*

3.4 A poset of loose transverse maps

In order to show the contractibility of $|\mathcal{T}|_{\mathrm{poly}}$ we introduce a poset $\mathcal{L} = \mathcal{L}(\Sigma, f)$ of *loose* transverse maps (see Definition 3.3) with exactly two transversion points on every cycle of $\bigvee^r S^1$. This poset gives rise to a subdivision of the polysimplicial complex $|\mathcal{T}|_{\mathrm{poly}}$, which is well-adapted to the surgeries we perform to prove contractibility. We are not able to prove contractibility directly with the constructions $|\mathcal{T}|_{\mathrm{poly}}$ or $|\mathcal{T}|$ from Section 3.3. The relation between $\mathcal{L}$ and $\mathcal{T}$ is analogous to the relation between the set of matchings $\mathrm{MATCH}^{\kappa\equiv 1}$ appearing in Theorem 2.8 and the set of matchings $\overline{\mathrm{MATCH}}^*$ appearing in Theorem 2.9.

Definition 3.15. Let Σ, f and V_o be as in Definition 3.6. The *poset of loose bi-transverse maps realizing* (Σ, f), denoted $\mathcal{L} = \mathcal{L}(\Sigma, f)$ or $(\mathcal{L}, \preceq_{\mathcal{L}})$, consists of the set of isotopy classes relative to $\mathcal{L}{=}\mathcal{L}(\Sigma,f)$
V_o of ***loose*** transverse maps realizing (Σ, f) with parameters $\kappa_x = 1$ for all $x \in B$. The order is defined as follows: assume that g is a transverse map realizing (Σ, f) with $\kappa_x = \mathbf{3}$ for all $x \in B$, let h_1 be the transverse map obtained from g by forgetting the two exterior transversion points for every $x \in B$, and let h_2 be the one obtained from g by forgetting the two interior points for every $x \in B$. If h_1 and h_2 are loose, then $[h_1] \preceq_{\mathcal{L}} [h_2]$ in $\mathcal{L}$.

The geometric realization of $\mathcal{L}$, denoted $|\mathcal{L}|$, is the order complex of $\mathcal{L}$: the simplicial complex $|\mathcal{L}|$
with vertices corresponding to the elements of $\mathcal{L}$ and an m-simplex for every chain $[h_0] \prec \ldots \prec [h_m]$ of length $m + 1$.

In other words, $[h_1] \preceq_{\mathcal{L}} [h_2]$ whenever the x-arcs and curves of $[h_1]$ can be arranged to be "nested" inside those of h_2, i.e. to lie inside the $(x, 0)$-zones of h_2, for every letter $x \in B$, so that the resulting map is a legal transverse map with four transversion points for every x. Another way to put it is that $[h_1] \preceq_{\mathcal{L}} [h_2]$ if and only if there are representatives h_1' and h_2', respectively, which are identical as maps, and are transverse to four points $(x, 0), \ldots, (x, 3)$ in every cycle of $\bigvee^r S^1$, such that the "official" transversion points of h_1' are $(x, 1)$ and $(x, 2)$, while the "official" transversion points of h_2' are $(x, 0)$ and $(x, 3)$.

Note that the relation $\preceq_{\mathcal{L}}$ is indeed a partial order: as the transverse map h with $\kappa_x = 3$ for every $x \in B$ is allowed to be loose, i.e., to violate **Restriction 3**, we get the desired reflexivity: $[h] \preceq_{\mathcal{L}} [h]$ for every $[h] \in \mathcal{L}$. For transitivity, assume $[h_1] \preceq_{\mathcal{L}} [h_2] \preceq_{\mathcal{L}} [h_3]$. By definition, this means one can draw the x-arcs/curves of some $h_2' \in [h_2]$ inside the $(x,0)$-zones of h_3 to obtain a legal loose transverse map, and likewise to draw the x-arcs/curves of some $h_1' \in [h_1]$ inside the $(x,0)$-zones of h_2' to obtain a legal loose transverse map. The union of all three collections of arcs and curves gives a loose transverse map g with $\kappa_x = 5$ for all x. By forgetting $(x,1)$ and $(x,4)$ for every x, we get a map that shows $[h_1] \preceq_{\mathcal{L}} [h_3]$. Finally, if $[h_1] \preceq_{\mathcal{L}} [h_2] \preceq_{\mathcal{L}} [h_1]$, we obtain in a similar fashion a map g with $\kappa_x = 5$ in which the $(x,0)$-arcs/curves are isotopic to the $(x,2)$-arcs/curves. This forces the $(x,1)$-arcs/curves to be isotopic to $(x,0)$ and to $(x,2)$. Analogously, the $(x,4)$-collection is isotopic to the $(x,3)$-collection and to the $(x,5)$-collection. Thus $[h_2] = [h_1]$ and we have established antisymmetry.

Proposition 3.16. *The spaces $|\mathcal{T}|_{\text{poly}}$ and $|\mathcal{L}|$ are homeomorphic. Moreover, there exists an homeomorphism $\alpha : |\mathcal{L}| \xrightarrow{\cong} |\mathcal{T}|_{\text{poly}}$, through which the simplices of $|\mathcal{L}|$ subdivide the polysimplices of $|\mathcal{T}|_{\text{poly}}$.*

The proof relies on the following general lemma:

Lemma 3.17. *For a finite chain (totally ordered set) C, let $(P_C, \preceq)$ be the poset consisting of $\{(i,j) \in C \times C \,|\, i \leq_C j\}$ with partial order given by $(i_1,j_1) \preceq (i_2,j_2)$ if and only if $i_2 \leq_C i_1 \leq_C j_1 \leq_C j_2$. Then there is a canonical homeomorphism $f_C\colon |P_C| \to \overline{\Delta_C}$ from the order complex of P_C to the closed $(|C|-1)$-simplex with vertices the elements of C. Moreover, the family $\{f_C\}_C$ of homeomorphisms respects subsets: for every subset $C' \subseteq C$, $f_C\big|_{C'} = f_{C'}$ and in particular $f_C\left(|P_{C'}|\right) = \overline{\Delta_{C'}}$.*

Proof. We write points in $\overline{\Delta_C}$ as $\sum_{c\in C} t_c \cdot c$ with $t_c \geq 0$ and $\sum_c t_c = 1$. If $(i_0,j_0) \prec \ldots \prec (i_m,j_m)$ is a chain in P_C, we write a point in the corresponding m-simplex of $|P_C|$ as $t_0\cdot(i_0,j_0)+\ldots+t_m\cdot(i_m,j_m)$ with $t_\ell \geq 0$ for all $\ell \in [m]$ and $\sum t_\ell = 1$. However, we recursively define f_C on any linear combination of $(i_0,j_0), \ldots, (i_m,j_m)$ with image some linear combination of $\{c \in C\}$. The definition is the following:

$$f_C\left(\sum_{s=0}^{m} t_s \cdot (i_s,j_s)\right) \overset{\text{def}}{=} \sum_{s=0}^{m} \left(\frac{t_s}{2}\cdot i_s + \frac{t_s}{2}\cdot j_s\right).$$

We recursively define the converse map, $\phi_C\colon \overline{\Delta_C} \to |P_C|$, again on any linear combination. For $c \in C$:

$$\phi_C\left(t\cdot c\right) = t\cdot(c,c).$$

If $i_0 \leq_C i_1$ then

$$\phi_C\left(t_0\cdot i_0 + t_1\cdot i_1\right) = \begin{cases} (t_0-t_1)\cdot(i_0,i_0) + 2t_1\cdot(i_0,i_1) & t_0 > t_1 \\ 2t_0\cdot(i_0,i_1) & t_0 = t_1 \\ 2t_0\cdot(i_0,i_1) + (t_1-t_0)\cdot(i_1,i_1) & t_0 < t_1 \end{cases}.$$

Finally, if $i_0 \leq_C \leq i_1 \leq_C \ldots \leq_C i_m$, then

$$\begin{aligned}\phi_C\left(t_0\cdot i_0+\ldots+t_m\cdot i_m\right) &= \\ &= \begin{cases} \phi_C\left((t_0-t_m)\cdot i_0 + t_1\cdot i_1+\ldots+t_{m-1}\cdot i_{m-1}\right) + 2t_m\cdot(i_0,i_m) & t_0 > t_m \\ 2t_0\cdot(i_0,i_m) + \phi_C\left(t_1\cdot i_1+\ldots+t_{m-1}\cdot i_{m-1}\right) & t_0 = t_m \\ 2t_0\cdot(i_0,i_m) + \phi_C\left(t_1\cdot i_1+\ldots+t_{m-1}\cdot i_{m-1} + (t_m-t_0)\cdot i_m\right) & t_0 < t_m \end{cases}.\end{aligned}$$

It is easy to verify that f_C and ϕ_C are inverse to each other, that they are continuous and that, indeed, $f_C\big|_{C'} = f_{C'}$ for every subset $C' \subseteq C$. In Figure 3.3 we illustrate the resulting subdivision of $\overline{\Delta_C}$ when $C = \{0 \prec 1 \prec 2\}$. $\square$

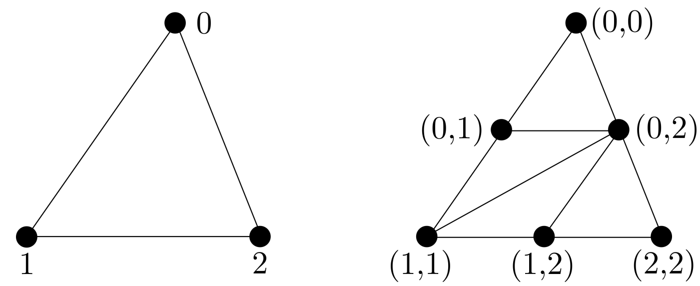

Figure 3.3: The subdivision (on the right) of the 2-simplex $\overline{\Delta_C}$ (on the left) given by the order complex of the poset of pairs P_C, where $C = \{0 \prec 1 \prec 2\}$ is the chain with three elements.

Proof of Proposition 3.16. Let $c = \{[h_0] \prec_{\mathcal{L}} \ldots \prec_{\mathcal{L}} [h_m]\}$ be a chain in $\mathcal{L}$. As above, find $h'_0, \ldots, h'_m$ so that $h'_j \in [h_j]$ and so that for every $0 \le j \le m-1$, the x-arcs/curves of h'_j are located inside the $(x,0)$-zones of h'_{j+1} and together they yield a legal (loose) transverse map with four transversion points for every letter. Let $g^{\mathrm{loose}}(c)$ be the loose transverse map with $2(m+1)$ transversion points for all $x \in B$, obtained as the union of the collections of arcs and curves of $h'_0, \ldots, h'_m$. Let $g^{\mathrm{strict}}(c)$ be the strict transverse maps obtained from $g^{\mathrm{loose}}(c)$ by forgetting every transverse point (x,j) such that the collection of (x,j)-zones of $g^{\mathrm{loose}}(c)$ violates **Restriction 3**, namely, so that the collection of (x,j)-arcs/curves is isotopic to the $(x,j+1)$-collection. Note that $\left[g^{\mathrm{strict}}(c)\right]$ is a well-defined element of $\mathcal{T}$, and we denote by $\{\kappa_x(c)\}_{x \in B}$ its parameters. For a singleton $[h] \in \mathcal{L}$, we denote also $g^{\mathrm{strict}}(h)$ the strict transverse map corresponding to the single-element chain $\{[h]\}$. g^{strict}

The sought-after homeomorphism $\alpha\colon |\mathcal{L}| \to |\mathcal{T}|_{\mathrm{poly}}$ is defined per simplex, where the simplex corresponding to the chain $c \subseteq \mathcal{L}$ is mapped into the polysimplex of $|\mathcal{T}|_{\mathrm{poly}}$ corresponding to $\left[g^{\mathrm{strict}}(c)\right]$. The exact definition goes through the single-letter case, using Fact 3.11. More concretely, for $[g] \in \mathcal{T}$, let $\mathcal{L}_{\le[g]} \overset{\mathrm{def}}{=} \left\{[h] \in \mathcal{L} \,\middle|\, \left[g^{\mathrm{strict}}(h)\right] \preceq_{\mathcal{T}} [g]\right\}$. While $\mathcal{L}$ is certainly *not* a product of its projections on the different letters $x \in B$, it is such a product locally inside $\mathcal{L}_{\le[g]}$: for $x \in B$, let $\gamma^x(g)$ denote the collection of x-arcs and curves of g (namely, the union over $j \in [\kappa_x]$ of (x,j)-arcs/curves). Let $P^x(g)$ denote the poset consisting of $\left\{\gamma^x_{i,j}(g)\right\}_{0 \le i \le j \le \kappa_x(g)}$, with $\gamma^x_{i_1,j_1}(g) \le_{P^x(g)} \gamma^x_{i_2,j_2}(g)$ if and only if $i_2 \le i_1 \le j_1 \le j_2$. Here, $\gamma^x_{i,j}(g)$ can be thought of as the union of (x,i)- and (x,j)-arcs/curves inside $\gamma^x(g)$. It is easy to see that $\mathcal{L}_{\le[g]}$ is isomorphic as a poset to a direct product of posets given by

$$\mathcal{L}_{\le[g]} \cong \prod_{x \in B} P^x(g),$$

where $[h] \in \mathcal{L}_{\le[g]}$ corresponds to $\prod_{x \in B} \gamma^x\left(g^{\mathrm{strict}}(h)\right)$. Hence,

$$\left|\mathcal{L}_{\le[g]}\right| = \left|\prod_{x \in B} P^x(g)\right| \overset{\text{Fact 3.11}}{\cong} \prod_{x \in B} |P^x(g)| \overset{\text{Lemma 3.17}}{\cong} \prod_{x \in B} \overline{\Delta_{\kappa_x(g)}} = \overline{\mathrm{polysim}}([g]),$$

and this homeomorphism defines $\alpha|_{\left|\mathcal{L}_{\le[g]}\right|}$. This definition expands to a well defined homeomorphism $\alpha\colon |\mathcal{L}| \to |\mathcal{T}|_{\mathrm{poly}}$ because for $[g'] \preceq_{\mathcal{T}} [g]$, the restriction of $\alpha_{\left|\mathcal{L}_{\le[g]}\right|}$ to $\left|\mathcal{L}_{\le[g']}\right|$ is exactly $\alpha_{\left|\mathcal{L}_{\le[g']}\right|}$. This shows that the image of the open simplices in $|\mathcal{L}|$ corresponding to the chains $\left\{c \,\middle|\, \left[g^{\mathrm{strict}}(c)\right] = [g]\right\}$ subdivides the open polysimplex $\mathrm{polysim}^\circ([g])$, and the image of $\left|\mathcal{L}_{\le[g]}\right|$ subdivides the closed simplex $\overline{\mathrm{polysim}}([g])$. □

3.5 Contractibility of the transverse map complex

To prove the contractibility of $|\mathcal{T}|_{\mathrm{poly}}$, we use *null-arcs*:

Definition 3.18. • A *null-arc* for (Σ, f) is an arc ω in Σ with endpoints in $V_o \subset \partial\Sigma$ and interior disjoint from $\partial\Sigma$, so that if ω is closed, it is not nullhomotopic, and such that $f_*(\omega) = 1$. The latter condition means, in other words, that the image of ω under f is nullhomotopic in $\bigvee^r S^1$ relative the endpoints. null-arc

- A *system of null-arcs* for (Σ, f) is a collection of null-arcs that are disjoint away from their endpoints and such that no two are isotopic relative to V_o.

- If Ω is a system of null-arcs for (Σ, f), then $\mathcal{T}_\Omega = \mathcal{T}_\Omega(\Sigma, f)$ and $\mathcal{L}_\Omega = \mathcal{L}_\Omega(\Sigma, f)$ are the subposets of $\mathcal{T}$ and $\mathcal{L}$, respectively, of isotopy classes of transverse maps which map $\bigcup_{\omega\in\Omega}\omega$ to $o \in \bigvee^r S^1$. $\mathcal{T}_\Omega, \mathcal{L}_\Omega$

Put differently, $\mathcal{T}_\Omega$ and $\mathcal{L}_\Omega$ consist of isotopy classes of transverse maps with arcs/curves collections that can be drawn away from Ω, meaning that every $\omega \in \Omega$ is entirely contained in some o-zone of the transverse map.

Note that $\mathcal{T}_\Omega$ and $\mathcal{L}_\Omega$ are downward closed: if $g' \preceq_{\mathcal{T}} g \in \mathcal{T}_\Omega$ then $g' \in \mathcal{T}_\Omega$ and likewise for $\mathcal{L}_\Omega$. Hence $|\mathcal{T}_\Omega|_{\mathrm{poly}}$ and $|\mathcal{L}_\Omega|$ are subcomplexes of $|\mathcal{T}|_{\mathrm{poly}}$ and $|\mathcal{L}|$, respectively. Moreover:

Claim 3.19. For any system of null-arcs for (Σ, f), the homeomorphism $\alpha\colon |\mathcal{L}| \to |\mathcal{T}|_{\mathrm{poly}}$ from Proposition 3.16 satisfies $\alpha(|\mathcal{L}_\Omega|) = |\mathcal{T}_\Omega|_{\mathrm{poly}}$.

Proof. The homeomorphism α maps the simplex corresponding to the chain c in $\mathcal{L}$ into the polysimplex corresponding to $\left[g^{\mathrm{strict}}(c)\right]$. But belonging to $\mathcal{T}_\Omega$ or to $\mathcal{L}_\Omega$ depends only on the o-zones of the transverse map, and the o-zones of the top element of c are identical to those of $g^{\mathrm{strict}}(c)$. Hence c is contained in $\mathcal{L}_\Omega$ if and only if its top element is in $\mathcal{L}_\Omega$, if and only if $\left[g^{\mathrm{strict}}(c)\right] \in \mathcal{T}_\Omega$. □

The following proposition is the main component of the proof of Theorem 3.14 concerning the contractibility of $|\mathcal{T}|_{\mathrm{poly}}$.

Proposition 3.20. *Let Ω be a system of null-arcs for (Σ, f), then there is a deformation retract of $|\mathcal{T}|_{\mathrm{poly}}$ to $|\mathcal{T}_\Omega|_{\mathrm{poly}}$. In particular, $\mathcal{T}_\Omega$ is non-empty.*

Proof of Theorem 3.14 assuming Proposition 3.20. Since the number of non-isotopic null-arcs that coexist for (Σ, f) is bounded by Euler characteristic considerations, it is obvious there exist *maximal* systems of null-arcs: systems so that no further null-arcs can be added to. Let Ω be a maximal system of null-arcs. We claim that $|\mathcal{T}_\Omega|_{\mathrm{poly}}$ is a single vertex of $|\mathcal{T}|_{\mathrm{poly}}$. This is enough by Proposition 3.20.

By Proposition 3.20, $\mathcal{T}_\Omega$ is non-empty. Since $\mathcal{T}_\Omega$ is downward closed, we can choose $g \in \mathcal{T}_\Omega$ with parameters $\kappa_x = 0$ for all x. Showing that $|\mathcal{T}_\Omega|_{\mathrm{poly}}$ is a single vertex is equivalent to showing that g is the only point in $\mathcal{T}_\Omega$.

To proceed, we claim that every connected component of $\Sigma \setminus \Omega$ has one of the following forms (and see Figure 3.4):

(i) **A rectangle around some arc β of g** This usually means a rectangle cut out by two null-arcs which are parallel to β with endpoints at the points of V_o neighboring the endpoints of β. But we also refer here to a bigon cut out by a single null-arc if β connects two adjacent components of $\partial_i\Sigma \setminus V_o$, which is possible when the word w_i is not cyclically reduced.

(ii) **A triangle bounded by three null-arcs**

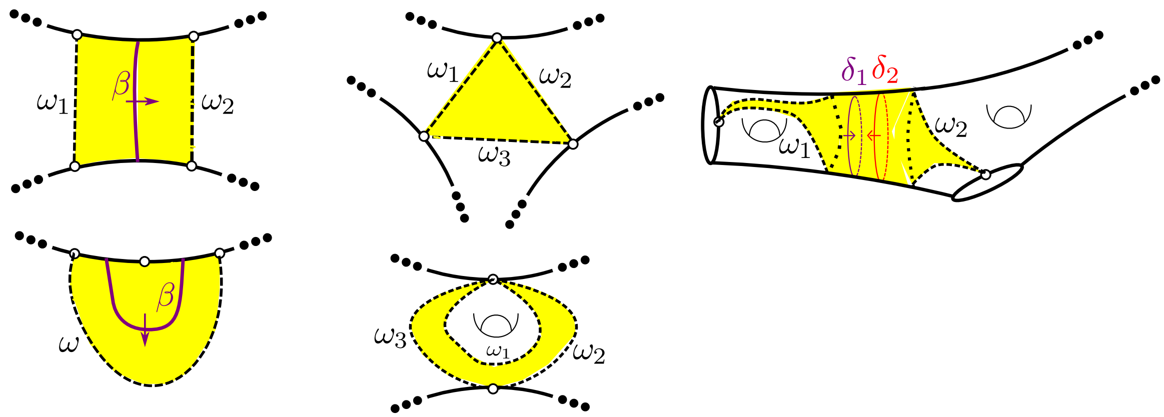

Figure 3.4: This figure shows the different types of connected components of $\Sigma \setminus \Omega$, where Ω is a maximal system of null arcs. In the terminology of the proof of Theorem 3.14 on Page 32, the two drawings on the left are pieces of type (i), namely, pieces containing a single arc β of the unique transverse map $g \in \mathcal{T}_\Omega$. The two drawings in the middle are pieces of type (ii): triangles bounded by three null-arcs. The drawing on the right is a piece of type (iii): an annulus cut out by two closed null-arcs and containing at least one curve of g (in the drawing: two curves, δ_1 and δ_2, corresponding to two different basis elements).

(iii) **An annulus cut out by two closed null-arcs** In this case the annulus must contain at least one curve of g (non-nullhomotopic, evidently).

Indeed, it is clear that for any arc β of g, the arc that is parallel to β on either side with endpoints at the points of V_o neighboring the endpoints of β is a null-arc and therefore in Ω by maximality. So every connected component of $\Sigma \setminus \Omega$ that contains an arc of g, contains a single arc of g and is of type (i). This also shows that components of type (i) touch all of $\partial\Sigma$. Any other component Σ' of $\Sigma \setminus \Omega$ does not contain any arc from g and does not touch $\partial\Sigma \setminus V_o$. If Σ' contains no curves of g neither, it can be triangulated by null-arcs and therefore has to be a triangle as in (iii) by the maximality of Ω.

Finally, assume that Σ' contains a curve δ of g. First, any component of $\partial\Sigma'$ is a chain of null-arcs, and by maximality has to consist of a single closed null-arc. Recall that Σ' contains no arcs of g, and that any non-nullhomotopic simple closed curve $c \subset \Sigma'$ disjoint from the curves of g is a null-curve (see Definition 1.5). If Σ' is not as described in item (iii), then one can add a null-arc to Ω inside Σ' in one of the following ways: if Σ' has at least two boundary components, draw a curve which leaves the marked point at one boundary component ω_1, takes some path to a different boundary component ω_2, goes around the ω_2 and returns to ω_1 along the same way; If Σ' has only one boundary component ω_1, there must be a pair of pants contained in Σ' which is free from curves of g, and one can draw a new null-curve by going from the marked point of ω_1, entering the pair of pants through one sleeve, circling another sleeve and going back. This is a contradiction to maximality. Hence Σ' is necessarily of type (iii).

We can now finish the argument showing that $[g]$ is the only element in $\mathcal{T}_\Omega$. Let g' be a transverse map for (Σ, f) with $[g'] \in \mathcal{T}_\Omega$. Obviously, there are no arcs/curves of g' in components of $\Sigma \setminus \Omega$ of type (ii). Any z-zone of g' is contained in some component Σ' of type (i) or (iii). But the structure of these components guarantees that any such z-zone is either a rectangle or an annulus. Thus $\kappa_x(g) = 0$ for all $x \in B$, for otherwise g' violates **Restriction 3**. We can now see that $[g'] = [g]$: its clear that their arcs are isotopic by the structure of type-(i) components. Their curves are also isotopic because for every Σ' of type (iii), consider an arc α connecting the two distinct marked points from V_o touching Σ'. The image of α under f completely prescribes the curves of g' inside Σ' (here we use also **Restriction 1**). □

3.5.1 Proof of Proposition 3.20

Now we come to prove Proposition 3.20 and show that $|\mathcal{T}|_{\text{poly}}$ deformation retracts to $|\mathcal{T}_\Omega|_{\text{poly}}$. Using Proposition 3.16 and Claim 3.19, we actually prove the equivalent statement that $|\mathcal{L}|$ deformation retracts to $|\mathcal{L}_\Omega|$. The general strategy to prove Proposition 3.20 is to perform local surgeries to gradually simplify transverse maps by removing intersections of their arcs and curves with the null-arcs in Ω. The complexity of a given transverse map in $\mathcal{L}$ is measured in terms of "depth of words along null-arcs":

Depth of words along null-arcs

Fix an arbitrary orientation along every null-arc in Ω. For every element $[h'] \in \mathcal{L}$, pick a loose transverse map $h \in [h']$ so that the arcs and curves are in minimal position with respect to Ω, meaning there are no bigons cut out by Ω and the arcs/curves of h. Every null-arc $\omega \in \Omega$ may cross arcs and curves of h, and we record these crossings as a word $u_\omega(h)$, writing $u_\omega(h)$

$$\begin{aligned} &P_x \text{ if the arc/curve has color } (x,0),\\ &Q_x \text{ if the arc/curve has color } (x,1). \end{aligned}$$

Put differently, we consider the path $h(\omega)$ in $\bigvee^r S^1$, and write P_x whenever it crosses $(x,0)$ and Q_x whenever it crosses $(x,1)$. Note that $h(\omega)$ begins and ends at o, and as ω is a null-arc and h homotopic to f, we get that $h(w)$ is nullhomotopic relative to its endpoints. This means that the word $u_\omega(h)$ can be reduced to the empty word by repeatedly deleting consecutive pairs of the form P_xP_x or Q_xQ_x.

For a general word in the alphabet $\{P_x, Q_x\}_{x\in B}$, we define its length as the length of its reduced form (it is standard the the reduced form does not depend on the choice of series of reduction steps). We define *the depth* of a word as the maximal length of a prefix. For example, in the word below, which reduces to the empty word, the superscripts denote the length of each prefix:

$${}^0P_x{}^1Q_x{}^2P_y{}^3P_y{}^2Q_z{}^3Q_z{}^2Q_t{}^3P_t{}^4Q_t{}^5Q_t{}^4P_t{}^3Q_t{}^2Q_x{}^1P_x{}^0.$$

Hence the depth of this word is 5. We denote the depth of the word $u_\omega(h)$ by $\operatorname{depth}(u_\omega(h))$. $\operatorname{depth}(u_\omega(h))$

Notice that $\operatorname{depth}(u_\omega(h)) = 0$ if and only if ω does not intersect any arcs or curves of h, namely, if and only if ω is contained inside some o-zone of h. Thus $[h] \in \mathcal{L}_\Omega$ if and only if $\operatorname{depth}(u_\omega(h)) = 0$ for all $\omega \in \Omega$.

We use the depth to filter $\mathcal{L}$: for $n \in \mathbb{Z}_{\geq 0}$ we let

$$\mathcal{P}_n \stackrel{\text{def}}{=} \{[h] \in \mathcal{L} \,|\, \operatorname{depth}(u_\omega(h)) \leq n \text{ for all } \omega \in \Omega\}.$$

Then

$$\mathcal{L}_\Omega = \mathcal{P}_0 \subseteq \mathcal{P}_1 \subseteq \ldots \subseteq \mathcal{P}_n \subseteq \ldots \subseteq \mathcal{L}$$

is a countable filtration of $\mathcal{L}$ and

$$\bigcup_{n=0}^{\infty} \mathcal{P}_n = \mathcal{L}.$$

A deformation retract $|\mathcal{P}_n| \to |\mathcal{P}_{n-1}|$

Let h with $[h] \in \mathcal{L}$ and $\omega \in \Omega$ satisfy that $\operatorname{depth}(u_\omega(h)) = n$, and consider the prefixes of length n in $u_\omega(h)$. If the last letter of such a prefix is, say, P_x, then so is the following letter. Each of these two letters correspond to a point where ω crosses an $(x,0)$-arc/curve of h. We call the segment of ω cut out by these two crossing points a *depth-n leaf* of h in Ω. The deformation retract we shall construct "prunes" all depth-n leaves of the elements of $\mathcal{P}_n$. depth-n leaf

Parity assumption A crucial observation here is that for every null arc ω and every h, if we cut ω to segments using the crossing points with the arcs and curves of h, then the segments alternate between belonging to o-zones of h and belonging to z-zones of h, with the first segment always in an o-zone. So if n is even, every depth-n leaf is contained in some o-zone, while if n is odd, every depth-n leaf is contained in some z-zone. In what follows we assume that n is even and so all depth-n leaves are contained in o-zones. The other case is very similar, and we shall point out steps of the proof where there is an important difference between the two cases.

The deformation retract $|\mathcal{P}_n| \to |\mathcal{P}_{n-1}|$ is based on a map $r_n \colon \mathcal{P}_n \to \mathcal{P}_{n-1}$ between the underlying posets.

Definition 3.21. For $[h] \in \mathcal{P}_n$ assume that h is in minimal position with respect to Ω. Define $r_n([h])$ by the following two steps: r_n
(i) Perform an H-move (see Definition 3.5) along every depth-n leaf of h in Ω to obtain h', a transverse map for (Σ, f).
(ii) If n is even (respectively, odd) consider all o-zones (respectively, z-zones) in h' which violate **Restriction 1** and remove them[14] to obtain h'', a transverse map for (Σ, f). Then set $r_n([h]) \overset{\text{def}}{=} [h'']$.

Recall that all null-arcs in Ω are disjoint away from their endpoints, so all depth-n leaves of h are disjoint, and so the different H-moves in step (i) do not interact with each other and can be performed simultaneously. Also note that $r_n([h])$ does not depend on the representative h of $[h]$. See Figure 3.5 for an illustration of how the r_n act on transverse maps.

We still need to explain why $r_n([h]) \in \mathcal{P}_{n-1}$. We do this through the following series of claims:

Claim 3.22. $\kappa_x(h'') = 1$ for all $x \in B$.

Proof. It is clear that step (i) of Definition 3.21 does not alter κ_x, so $\kappa_x(h') = 1$. It remains to show that for every $x \in B$ and $j \in [1]$, some (x,j)-arc/curve survives step (ii). We remark that this is clear if there is some (x,j)-arc in h, because r_n does not modify h near $\partial\Sigma$. It is less clear, however, when there are only (x,j)-curves.

Let β be some (x,j)-arc or (x,j)-curve of h, and consider O_1, the o-zone of h touching β. All depth-n leaves of h are contained inside o-zones (recall our ongoing assumption in the proofs that n is even), and the leaves inside O_1 cut it in step (i) to smaller o-zones of h', separated by "z-tunnels" along the leaves of depth n. Let $\mathcal{O}_{(x,j)}$ denote the collection of o-zones of h' which are contained in O_1 and which are removed in step (ii) because they contain no marked points and have only (x,j)-arcs/curves along their boundary. If $\mathcal{O}_{(x,j)}$ is empty, we are done, as the (x,j)-arcs/curves which are the traces of β survive in h''. So assume $\mathcal{O}_{(x,j)}$ is non-empty. It cannot include all the o-zones of h' contained in O_1, because this would mean that O_1 itself is redundant. Thus, there must be some o-zone $O_2 \in \mathcal{O}_{(x,j)}$ which borders, through a depth-n leaf, some o-zone $O_3 \notin \mathcal{O}_{(x,j)}$ of h' which is contained in O_1. Since the leaf separating O_2 and O_3 has (x,j)-arcs/curves on both sides (in h'), O_3 has some bounding (x,j)-arc/curve, which survives step (ii). ☐

We have not shown yet that $r_n([h]) \in \mathcal{L}$: it remains to prove that h'' is loose, but the following claim is the analog of saying that $[h] \preceq_{\mathcal{L}} [h'']$:

Claim 3.23. There is a transverse map g for (Σ, f) with $\kappa_x = 3$ for all x, so that forgetting $(x,0)$ and $(x,3)$ for all x yields a map in $[h]$ and forgetting $(x,1)$ and $(x,2)$ for all x yields a map in[15] $[h'']$.

---

[14] As we explained in the proof of Lemma 3.8, in the current scenario, a zone violating **Restriction 1** is necessarily a zone bounded by curves all of which are of the same color. By removing the zone we mean removing all bounding curves to obtain a new transverse map, and this procedure does not change the homotopy type of the map relative to V_o.

[15] If n is odd, the parallel claim is the analog of $[h''] \preceq_{\mathcal{L}} [h]$.

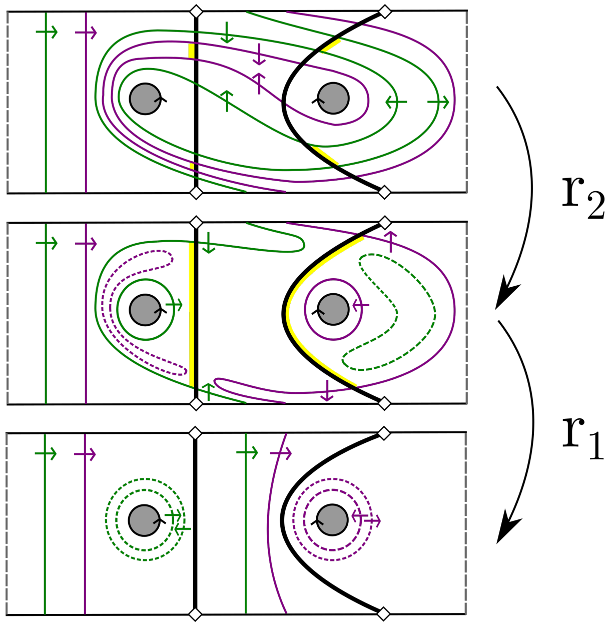

Figure 3.5: This figure shows the effects of r_2 and r_1 on a transverse map on a genus 1 surface with 2 boundary components. The surface is depicted as a rectangle with 2 holes (shaded) whose boundaries are identified according to the labeled orientations, and with the two dashed vertical sides of the rectangle also identified. Green corresponds to $(x, 0)$ and purple corresponds to $(x, 1)$. The two null-arcs in the system are the thick black arcs. Yellow shading indicates depth-n leaves of the guide arcs. Dashed curves are those to be removed by Step (ii) of r_i – see Definition 3.21.

Proof. First, construct a transverse map g with $\kappa_x = 3$ by duplicating the arcs and curves of h, so that the $(x, 0)$-arcs/curves are isotopic to the $(x, 1)$-arcs/curves, and likewise with $(x, 2)$ isotopic to $(x, 3)$. Since the H-moves in step (i) of Definition 3.21 are performed in o-zones (recall our assumption that n is even), we can perform them for g, in which they involve only arcs/curves with color from $\bigcup_x \{(x, 0), (x, 3)\}$ and occur inside o-zones. The resulting map, call it g', shows the analog of $h \preceq_{\mathcal{L}} h'$. Finally, the o-zones of g' are identical (up to homotopy) to those of h', so step (ii) can be performed in g' by removing all redundant o-zones of g'. The resulting map, g'', is still transverse with parameters $\kappa_x = 3$ for all x by Claim 3.22, and is the map we need to establish the claim. $\square$

Lemma 3.24. *$r_n([h]) \in \mathcal{L}$.*

Proof. We need to show that h'' is loose, namely that it abides to **Restrictions 1** and **2**. Neither step (i) nor step (ii) from Definition 3.21 change h near $\partial\Sigma$, so h'' abides to **Restriction 2** because so does h. It remains to show there are no "redundant" zones in h'', namely, no zones which violate **Restriction 1**. Note that the removal of redundant o-zones of h' in step (ii) enlarges z-zones and possibly merges several z-zones into one, but it does not create new o-zones nor does it affect other existing o-zones. So the remaining o-zones are not redundant.

As for z-zones, we use the map g'' from Claim 3.23. We claim that g'' has no redundant z-zones. Clearly, g'' has no $(x, 1)$-redundant zone, because these are exactly the $(x, 0)$-zones of h, which abide to **Restriction 1**. Note that every $(x, 0)$-arc/curve of g'' is parallel, at least in some segments, to $(x, 1)$-arcs/curves (by the nature of H-moves). Hence, every $(x, 0)$-zone of g'' must have some bounding $(x, 1)$-arc/curve. Therefore, a redundant $(x, 0)$-zone in g'' has only $(x, 1)$-arcs/curves at its boundary, and is thus a redundant o-zone of h, a contradiction. That there are no redundant $(x, 2)$-zones in g'' is analogous to the $(x, 0)$ case.

Now, let Z be an arbitrary z-zone of h''. Without loss of generality, there is some $(x,0)$-arc/curve of h'' at ∂Z. This $(x,0)$-arc/curve is at ∂Z_0 for some $(x,0)$-zone Z_0 of g contained in Z. By the claim on g, this Z_0 borders some $(x,1)$-zone $Z_1 \subset Z$ of g, which borders some $(x,2)$-zone $Z_2 \subset Z$ of g. But Z_2 has some $(x,3)$-arc/curve of g at its boundary, which is necessarily a $(x,1)$-arc/curve of h'' at the boundary of Z. Hence Z does not violate **Restriction 1**. $\square$

Corollary 3.25. *$r_n([h]) \in \mathcal{P}_{n-1}$ and*[16] *$[h] \preceq_{\mathcal{L}} r_n([h])$.*

Proof. It remains to show that $\operatorname{depth}(u_\omega(h'')) \leq n-1$ for all $\omega \in \Omega$. The H-moves of step (i) in the definition of r_n remove all the crossings between ω and arcs/curves of h which cut out depth-n leaves. It is thus clear that $\operatorname{depth}(u_\omega(h')) \leq n-1$. But whenever ω enters a redundant zone of h', it has to leave it through an arc/curve of the same color. So the effect of removing a redundant zone on the words $u_w(h')$ is performing reduction steps (omitting consecutive pairs of the type P_xP_x or Q_xQ_x). Reduction moves cannot increase the depth of the word. $\square$

After establishing that $r_n\colon \mathcal{P}_n \to \mathcal{P}_{n-1}$, our next goal is to use r_n to obtain the sought-after deformation retract. We do this using the following general technique concerning posets:

A map $\varphi\colon P \to Q$ between posets which is order-preserving, in the sense that $p_1 \leq_P p_2 \implies \varphi(p_1) \leq_Q \varphi(p_2)$, maps a chain $p_0 <_P \ldots <_P p_m$ in P to a, possibly "stuttering", chain $\varphi(p_0) \leq_Q \ldots \leq_Q \varphi(p_m)$ in Q, so the set $\{\varphi(p_0), \ldots, \varphi(p_m)\}$ defines a simplex in the order complex $|Q|$. This allows the following natural induced map $|\varphi| : |P| \to |Q|$ between the order complexes:

$$|\varphi| \left(\sum \lambda_i p_i\right) = \sum \lambda_i \varphi(p_i). \tag{3.2}$$

Lemma 3.26. *Let P be a subposet of the poset Q. Assume that $\varphi\colon Q \to P$ satisfies the following three conditions:*

- *φ is order-preserving*
- *φ is a retract, i.e. $f\big|_P \equiv \mathrm{id}$*
- *$\varphi(q) \leq q$ for all $q \in Q$, or $\varphi(q) \geq q$ for all $q \in Q$*

Then $|\varphi| : |Q| \to |P|$ is a strong deformation retract.

By a strong deformation retract we mean that there is a homotopy of $|\varphi|$ with the identity on $|Q|$ which fixes $|P|$ pointwise throughout the homotopy.

Proof. Recall that a map ψ between posets is called a poset-morphism if it is order preserving. If $\psi\colon P \to Q$ is a poset morphism, we let $|\psi|$ denote the induced map $|\psi| : |P| \to |Q|$ defined as in (3.2). If P and Q are posets, $\psi_0, \psi_1\colon P \to Q$ are poset morphisms, and $\psi_0(p) \leq \psi_1(p)$ for every $p \in P$, then $|\psi_0|$ and $|\psi_1|$ are homotopic. Indeed, let $\{0 \leq 1\}$ denote the poset with two comparable elements 0 and 1. Define a map $(\psi_0, \psi_1) : P \times \{0 \leq 1\} \to Q$ by $(p,0) \mapsto \psi_0(p)$ and $(p,1) \mapsto \psi_1(p)$. This is clearly a poset-morphism by the assumptions, so it induces a continuous map

$$|(\psi_0, \psi_1)| : |P \times \{0 \leq 1\}| \to |Q|.$$

By Fact 3.11, there is an homeomorphism

$$|P \times \{0 \leq 1\}| \xrightarrow{\cong} |P| \times |\{0 \leq 1\}| = |P| \times [0,1],$$

[16] For n odd, $r_n([h]) \preceq_{\mathcal{L}} [h]$.

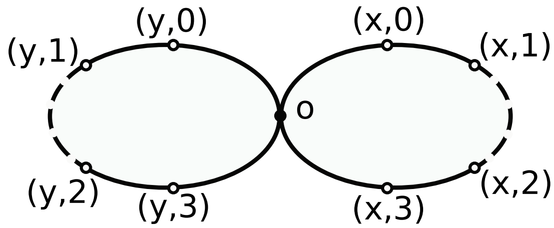

Figure 3.6: The wedge $\bigvee^r S^1$ when $r = 2$, $B = \{x, y\}$. The marked points are the transversion points of a map g, with $\kappa_x(g) = \kappa_y(g) = 3$. If h_1 is the transverse map obtained from g by forgetting $(x, 0), (x, 3), (y, 0), (y, 3)$ and if n is even, then a depth-n leaf γ of h_1 at some null-arc ω is contained in some o-zone of h_1, meaning that $g(\gamma)$ lives outside the broken segments in the figure.

so we get that $|(\psi_0, \psi_1)|$ is a continuous map $|P| \times [0, 1] \to |Q|$. Because $|(\psi_0, \psi_1)|\Big|_{|P\times\{0\}|} \equiv |\psi_0|$ and $|(\psi_0, \psi_1)|\Big|_{|P\times\{1\}|} \equiv |\psi_1|$, the map $|(\psi_0, \psi_1)|$ is the sought-after homotopy. (This result appears in [Qui73, Section 1.3].)

Note that the map $\varphi\colon Q \to Q$ in the statement of the lemma and the identity $\mathrm{id}\colon Q \to Q$ satisfy the conditions regarding ψ_0 and ψ_1 above. Hence $|\varphi|$ is homotopic to the identity. The fact that the homotopy fixes $|P|$ pointwise follows from the fact that the homotopy above does not move the points where ψ_0 and ψ_1 agree. Namely, if $P_0 \subseteq P$ is the subposet where $\psi_0(p) = \psi_1(p)$, then $|(\psi_0, \psi_1)|(x, t) = \psi_0(x) = \psi_1(x)$ for every $x \in |P_0|$ and $t \in [0, 1]$. $\square$

Proposition 3.27. *The map $r_n\colon \mathcal{P}_n \to \mathcal{P}_{n-1}$ satisfies the conditions of Lemma 3.26 and so defines a strong deformation retract*

$$|r_n| : |\mathcal{P}_n| \to |\mathcal{P}_{n-1}|.$$

Proof. We already proved above that for n even, $[h] \preceq_{\mathcal{L}} r_n([h])$ which is the third assumption of Lemma 3.26. The second assumption is also clear: if $[h] \in \mathcal{P}_{n-1}$, then h admits no depth-n leaves in Ω, and therefore in Definition 3.21, $h = h' = h''$. It remains to show that r_n is order-preserving.

Let h_1 and h_2 be transverse maps so that $[h_1], [h_2] \in \mathcal{P}_n$, with $[h_1] \preceq_{\mathcal{L}} [h_2]$, and assume that g is a transverse map g with $\kappa_x = 3$ for all x, so that h_1 and h_2 are obtained by forgetting the exterior and interior, respectively, two transversion points for every letter x. We also assume g is in minimal position with respect to Ω. For $\omega \in \Omega$, the words $u_\omega(h_1)$ and $u_\omega(h_2)$ are very much dependent: they can be constructed simultaneously by following the path $g(\omega)$ in $\bigvee^r S^1$, and adding a letter to $u_\omega(h_1)$ whenever $g(\omega)$ crosses some $(x, 1)$ or $(x, 2)$ point, and a letter to $u_\omega(h_2)$ whenever $g(\omega)$ crosses some $(x, 0)$ or $(x, 3)$ point. This description shows that whenever ω visits an o-zone or an $(x, 1)$-zone of g, the prefix of the two words until that point has the same reduced form and, in particular, the same length.

Consider a depth-n leaf γ of h_1 in ω. The beginning of γ is at a crossing point of ω with some (x, j)-arc/curve of g with $x \in B$ and $j \in \{1, 2\}$, in which ω leaves an $(x, 1)$-zone of g and enters some $(x, 0)$- or $(x, 2)$-zone. Without loss of generality, assume that ω crosses some $(y, 1)$-arc/curve with $y \in B$. The image $g(\gamma)$ is a closed path in $\bigvee^r S^1$, based at $(y, 1)$, which avoids the segments $[(x, 1), (x, 2)]$ for every $x \in B$ – see Figure 3.6.

When one follows the prefix of the word $u_\omega(h_2)$ along γ, it is clear, therefore, that at the beginning of γ it has length $n - 1$. If it then crosses $(y, 0)$, it has the same length as the prefix of $u_\omega(h_1)$ which is n. Then, it could seemingly cross, e.g., $(x, 3)$ for some $x \in B$, but this would increase the length of the prefix of $u_\omega(h_2)$ to $n+1$, which is impossible as $[h_2] \in \mathcal{P}_n$. Hence $g(\gamma)$ can only cross the point $(y, 0)$ back and forth. Every two consecutive such crossings define a depth-n leaf of h_2 at ω.

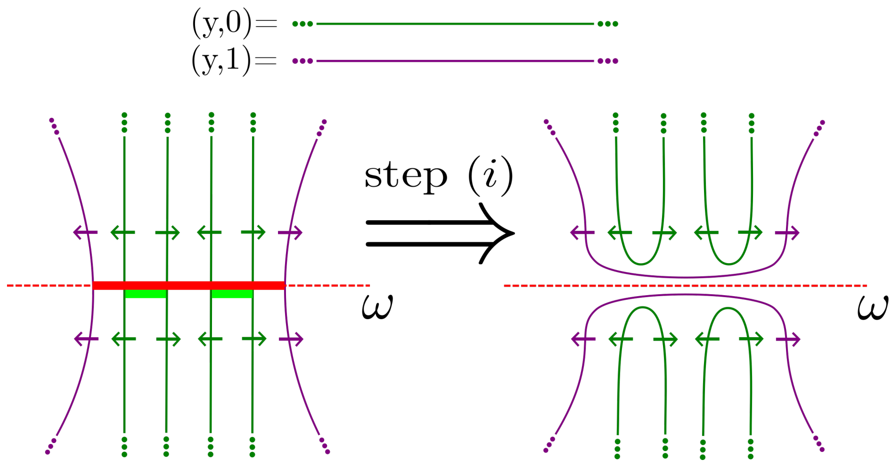

Figure 3.7: On the left, a piece of a null-arc ω crosses some arcs/curves of g, the transverse map with $\kappa_x = 3$ for all x showing that $h_1 \preceq_{\mathcal{L}} h_2$, both inside $\mathcal{P}_n$. The thick red part of ω is a depth-n leaf γ of h_1, which in g is a segment of ω between two crossing-points with $(y,1)$-arcs/curves. Inside γ there are two depth-n leaves of h_2, which, in g, are two segments cut out by $(y,0)$-arcs/curves. The right hand side shows the result of step (i) of Definition 3.21 on this local picture, where performing the single H-move for h_1 after the two H-moves of h_2 causes no collisions.

Therefore, in the notation of Definition 3.21, step (i) can be performed in two phases: first, perform step (i) for h_2, where the depth-n leaves never cross any arcs/curves of h_1. Second, perform step (i) for h_1: although a depth-n leaf of h_1 may cross arcs/curves of h_2, the previous paragraph explains why it never crosses arcs/curves of h_2'. The resulting h_1' and h_2' are compatible together in the sense there is g' with $\kappa_x = 3$ as in the definition of the order on $\mathcal{L}$ (although, of course, h_1' and h_2' may not be in $\mathcal{L}$). See Figure 3.7.

In step (ii) of Definition 3.21 we now remove redundant o-zones of h_1' and of h_2'. Since the o-zones of h_2' and those of g' coincide, removing redundant o-zones of h_2' is equivalent to removing redundant o-zones of g' and keeps the structure of g' as a legal transverse map with $\kappa_x = 3$ for all x. Denote the resulting map by $\overline{g'}$. However, we still need to show that removing redundant o-zones of h_1' does not cause a problem, namely, that any redundant o-zone of h_1' does not contain any curves of h_2'', which are the same as $(x,0)$- or $(x,3)$-curves of $\overline{g'}$ for any $x \in B$.

Indeed, let O be a redundant o-zone of h_1'. Without loss of generality it is bounded by outgoing $(y,1)$-curves of $\overline{g'}$. Assume there is some curve of h_2'' inside O which is not a $(y,0)$-curve of $\overline{g'}$, say, an $(x,3)$-curve of $\overline{g'}$. The negative side of this $(x,3)$-curve cannot be a redundant $(x,2)$-zone of $\overline{g'}$, because then it would be a redundant z-zone of h_2'' which is impossible by Lemma 3.24. Thus, this $(x,2)$-zone of $\overline{g'}$ must have some $(x,2)$-arc/curve at its boundary, a contradiction to the assumption that O is redundant. We conclude that O may only contain $(y,0)$-curves of $\overline{g'}$. But then, on their negative side, these curves must bound a redundant zone (there cannot be marked points from V_o inside as O is redundant), and thus should have been removed in step (ii) for h_2. Therefore, step (ii) for h_1' can be performed on $\overline{g'}$ without violating any rule, and the resulting map, g'', shows that $r_n(h_1) = [h_1''] \preceq_{\mathcal{L}} [h_2''] = r_n([h_2])$. □

Proof of Proposition 3.20. To get a deformation retract of $|\mathcal{L}|$ to $|\mathcal{L}_\Omega|$ we perform $|r_n|$ at time $\left[\frac{1}{2^n}, \frac{1}{2^{n-1}}\right]$. We remark that the fact that $|r_n|$ is a *strong* deformation retract, namely, keeps $|\mathcal{P}_{n-1}|$ fixed pointwise, guarantees that the total deformation retract on $|\mathcal{L}|$ is well defined. □

4 The action of $\mathrm{MCG}(f)$ on the complex of transverse maps

In this section we prove our main results: Theorems 1.4, 1.6 and 1.7. We begin with some background on L^2-Euler characteristics.

4.1 L^2-Betti numbers and L^2-Euler characteristics

We now define the L^2-invariants of groups that appear in our main theorem, although, for the sake of the proofs, one can use Theorem 4.2, Lemma 4.3 and Theorem 4.4 as black boxes.

The following definitions and properties are all found in the book of Lück [Lüc02]; many of the ideas we discuss originate from the paper of Cheeger and Gromov [CG86]. Throughout this subsection, G is a discrete group.

Definition 4.1 ([Lüc02, Def. 1.25])**.** A G-CW-complex is a CW-complex with a cellular action of G such that if an element of G fixes an open cell, it acts as the identity on that open cell.

Following [Lüc02, Def. 1.1], the *group von Neumann algebra* $\mathcal{N}(G)$ is defined to be the space of G-equivariant bounded operators from $\ell^2(G)$ to itself. Here $\ell^2(G)$ is given the standard Hermitian inner product making it a Hilbert space. Now suppose X is a G-CW-complex. Denote by $C_*^{\text{sing}}(X)$ the singular chain complex of X. This is a complex of left $\mathbb{Z}G$-modules. Giving $\mathcal{N}(G)$ the structure of an $(\mathcal{N}(G), \mathbb{Z}G)$-bimodule, we can form a chain complex

$$\dots \xrightarrow{d_{p+1}} \mathcal{N}(G) \otimes_{\mathbb{Z}G} C_p^{\text{sing}}(X) \xrightarrow{d_p} \mathcal{N}(G) \otimes_{\mathbb{Z}G} C_{p-1}^{\text{sing}}(X) \xrightarrow{d_{p-1}} \dots$$

of $\mathcal{N}(G)$-modules. This is a Hilbert chain complex in the terminology of [Lüc02, Def. 1.15]. In particular, each piece $\mathcal{N}(G) \otimes_{\mathbb{Z}G} C_p^{\text{sing}}(X)$ is a Hilbert module for $\mathcal{N}(G)$ as defined in [Lüc02, Def. 1.5], $\mathcal{N}(G) \otimes_{\mathbb{Z}G} C_p^{\text{sing}}(X)$ is a Hilbert space, and the boundary maps are bounded G-equivariant operators. The L^2-homology of the pair (X, G) we denote by $H_*^{(2)}(X; G)$ and define by

$$H_p^{(2)}(X; G) \overset{\text{def}}{=} \frac{\ker(d_p)}{\text{closure}(\text{image}(d_{p+1}))},$$

cf. [Lüc02, Def. 6.50, Def. 1.16]. Each of these homology groups are themselves Hilbert $\mathcal{N}(G)$-modules. Any $\mathcal{N}(G)$-module M has an associated dimension in $[0, \infty]$ called the *von Neumann dimension* and denoted by $\dim_{\mathcal{N}(G)}(M)$ [Lüc02, Def 6.20]. The L^2*-Betti numbers* of the pair (X, G) are defined by

$$b_p^{(2)}(X, G) \overset{\text{def}}{=} \dim_{\mathcal{N}(G)} H_p^{(2)}(X; G) \in [0, \infty].$$

If

$$\sum_{p \in \mathbb{Z}_{\geq 0}} b_p^{(2)}(X, G) < \infty \tag{4.1}$$

then we can also define the L^2*-Euler characteristic* of the pair (X, G) to be

$$\chi^{(2)}(X, G) = \sum_{p \in \mathbb{Z}_{\geq 0}} (-1)^p \cdot b_p^{(2)}(X, G) \in \mathbb{R}.$$

If EG is a contractible G-CW-complex with a free action of G then we define

$$b_p^{(2)}(G) \overset{\text{def}}{=} b_p^{(2)}(EG, G)$$

and if moreover (4.1) holds for $X = EG$, then we also define as in [Lüc02, Def. 6.79] the L^2*-Euler characteristic of* G to be

$$\chi^{(2)}(G) \overset{\text{def}}{=} \chi^{(2)}(EG, G).$$

Since EG is unique up to G-equivariant homotopy equivalence, it follows for example from [Lüc02, Theorem 6.54] that the quantities $b_p^{(2)}(G), \chi^{(2)}(G)$ only depend on G. The existence and G-homotopy uniqueness of EG is discussed in [Lüc02, pg. 33] with references therein to [tD72, tD87].

Assume X is an arbitrary G-CW-complex. If c is a cell of X write G_c for the isotropy group (stabilizer) of c in G. As in [Lüc02, §6.6.1], we consider the quantities

$$|G_c|^{-1}$$

where we set $|G_c|^{-1} = 0$ if G_c is infinite. We define following [Lüc02, Def 6.79]

$$m(X,G) := \sum_{[c]\in G\backslash X} |G_c|^{-1} \in [0,\infty].$$

Theorem 4.2 ([Lüc02, Thm. 6.80(1)])**.** *If $m(X,G)$ is finite then the sum of $b_p^{(2)}(X,G)$ is finite and, moreover,*

$$\chi^{(2)}(X,G) = \sum_{[c]\in G\backslash X} (-1)^{\dim c} |G_c|^{-1}. \tag{4.2}$$

Following [Lüc02, Def. 7.1] let $\mathcal{B}_\infty$ denote the class of groups G for which $b_p(G) = 0$ for all $p \in \mathbb{Z}_{\geq 0}$. $\mathcal{B}_\infty$

Lemma 4.3. *If X is a contractible G-*CW*-complex, and for all cells c of X the isotropy group G_c is either finite or in $\mathcal{B}_\infty$, then*

$$b_p^{(2)}(X,G) = b_p^{(2)}(G), \quad p \in \mathbb{Z}_{\geq 0}.$$

Hence if also $m(X,G)$ is finite then $\chi^{(2)}(X,G) = \chi^{(2)}(G)$.

Proof. This is [Lüc02, Exercise 6.20]. It can be proved by combining [Lüc02, Thm 6.54 (2) and (3)], and referring to Theorem 4.2 for the statement about Euler characteristics. □

To use Lemma 4.3 we need to have a source of groups lying in $\mathcal{B}_\infty$. The following theorem is essentially due to Cheeger and Gromov (cf. [CG86, Corollary 0.6]). The precise statement we need can be deduced from [Lüc02, Theorem 7.2, items (1) and (2)]. Recall that a discrete group is called *amenable* if it has a finitely additive left invariant probability measure.

Theorem 4.4 (Cheeger-Gromov)**.** *If G is a discrete group containing a normal infinite amenable subgroup then $G \in \mathcal{B}_\infty$.*

4.2 The complex of transverse maps as a $\mathrm{MCG}(f)$-CW-complex

The stabilizer $\mathrm{MCG}(f)$ of f in $\mathrm{MCG}(\Sigma)$ acts on the poset $\mathcal{T} = \mathcal{T}(\Sigma, f)$ by precomposition: if $[\rho] \in \mathrm{MCG}(f)$ and $[g] \in \mathcal{T}$ with parameters κ, then $[\rho].[g] = [g \circ \rho^{-1}]$ is an element of $\mathcal{T}$ with the same κ: indeed, $g \circ \rho^{-1}$ is a transverse map realizing f with the exact same transversion points as g. This action is obviously an order preserving action: if $[g_1] \preceq [g_2]$ then $[\rho].[g_1] \preceq [\rho].[g_2]$.

We now show that this action on $\mathcal{T}$ turns its geometric realization into a $\mathrm{MCG}(f)$-CW-complex, as in Definition 4.1. The properties mentioned above of the action of $\mathrm{MCG}(f)$ on the poset $\mathcal{T}$ guarantee that this is the case for $|\mathcal{T}|$, the order complex of $\mathcal{T}$ (see Page 28 for the definition of $|\mathcal{T}|$). We claim this is also the case for the polysimplicial complex $|\mathcal{T}|_{\mathrm{poly}}$:

Lemma 4.5. *Let $[(\Sigma, f)] \in \mathsf{Surfaces}(w_1, \ldots, w_\ell)$. Let $\Gamma = \mathrm{MCG}(f)$. The action of Γ on $\mathcal{T} = \mathcal{T}(\Sigma, f)$ makes $|\mathcal{T}|_{\mathrm{poly}}$ into a Γ-CW-complex.*

Proof. If $[\rho] \in \Gamma$ fixes $[g] \in \mathcal{T}$ we need to show $[\rho]$ cannot permute the faces of polysim ([g]). But $\left[g \circ \rho^{-1}\right] = [g]$ means there is an isotopy of transverse maps between g and $g \circ \rho^{-1}$. In such an isotopy, the $\sum_{x \in B} (\kappa_x(g) + 1)$ points of transversion in $\bigvee^r S^1$ may move around, but away from the wedge point o, and without collisions. This means that their order on each circle of $\bigvee^r S^1$ is preserved. In particular, for every $x \in B$ and $j \in [\kappa_x(g)]$, the isotopy takes the (x, j) point of g to the (x, j) point of $g \circ \rho^{-1}$, and the collection of (x, j)-arcs/curves of g to the collection of (x, j)-arcs/curves of $g \circ \rho^{-1}$. Thus, $[\rho]$ necessarily preserves every face of polysim $([g])$. □

Definition 4.6. We define $\mathcal{T}_\infty = \mathcal{T}_\infty(\Sigma, f)$ to be the subposet of $\mathcal{T} = \mathcal{T}(\Sigma, f)$ consisting of classes of transverse maps $[g]$ in $\mathcal{T}$ that do *not* fill Σ. $\mathcal{T}_\infty(\Sigma, f)$

Recall that $[g]$ fills Σ if its o-zones and z-zones are all topological discs. This means, in particular, that the preimage of every transversion point contains only arcs (and no curves).

Our notation $\mathcal{T}_\infty$ is in analogy to Harer's use of A_∞ in [Har86, Har85] for the subcomplex of the arc complex consisting of arc systems that do not cut the surface into discs; Harer used this complex in [Har86] to construct a Borel-Serre type bordification of Teichmüller space. This had previously been done by Harvey [Har81] using the complex of curves. These bordifications are closely related to the Deligne-Mumford compactification [DM69] of the moduli space of curves (see [Mon08, Remark 2.5]).

We define $|\mathcal{T}_\infty|_{\text{poly}}$ to be the polysimplicial subcomplex of $|\mathcal{T}|_{\text{poly}}$ consisting of polysimplices in $\mathcal{T}_\infty$. It is clear that $|\mathcal{T}_\infty|_{\text{poly}}$ is indeed a subcomplex of $|\mathcal{T}|_{\text{poly}}$ since if the arcs of $[g]$ do not cut Σ into discs then neither do the arcs of $[g']$ obtained from g by forgetting points of transversion.

Lemma 4.7. $\mathrm{MCG}(f)$ *acts freely on* $\mathcal{T} \setminus \mathcal{T}_\infty$.

Proof. If $[\rho]$ in $\mathrm{MCG}(f)$ fixes an isotopy class $[g]$ of filling transverse maps then we can assume ρ fixes all the arcs of g, so restricts to mapping classes on each of the zones of g, which are all discs. The Alexander Lemma [FM12, Lemma 2.1] implies these mapping classes must be trivial, so ρ is homotopic to the identity on each zone of g, hence overall. □

So the isotropy groups $\mathrm{MCG}(f)_{[g]}$ are trivial for $[g] \in \mathcal{T} \setminus \mathcal{T}_\infty$. The following lemma shows that for any other element of $\mathcal{T}$, the isotropy groups are not only infinite, but also have vanishing L^2-Betti numbers:

Lemma 4.8. *Let* $\Gamma = \mathrm{MCG}(f)$. *If* $[g] \in \mathcal{T}_\infty$ *then the isotropy group* $\Gamma_{[g]}$ *of* $[g]$ *are in* $\mathcal{B}_\infty$.

Proof. Fix a representative transverse map g for $[g]$. Let $\mathcal{C}$ denote a set of disjoint simple closed curves, where for every zone of g we add a simple closed curve parallel to every boundary component of that zone to $\mathcal{C}$, and we think of the curves as drawn inside the zones of g they come from. If g contains curves in the preimages of points of transversion, then this process can add to $\mathcal{C}$ multiple copies of isotopy classes of simple closed curves, but this does not matter.

Because $\mathcal{C}$ is drawn in the zones of g, then a Dehn twist in any element of $\mathcal{C}$ belongs to Γ. Let N be the subgroup of Γ generated by Dehn twists in elements of $\mathcal{C}$. This group is isomorphic to $\mathbf{Z}^r$ for some $r \geq 0$ because the curves in $\mathcal{C}$ are disjoint. In fact, $r \geq 1$ since by assumption, some zone of g is not a topological disc, and hence has a boundary component which does not bound a disc, so gives rise to a non-trivial Dehn twist. To see that N is normal in $\Gamma_{[g]}$, note that any mapping class in $\Gamma_{[g]}$ can be taken to permute the zones of g. Hence $\Gamma_{[g]}$ permutes the isotopy classes of curves in $\mathcal{C}$. Therefore the conjugation by $[\rho] \in \Gamma_{[g]}$ of any Dehn twist in an element of $\mathcal{C}$ is another Dehn twist in an element of $\mathcal{C}$.

It was proved by von Neumann [von29] that $\mathbb{Z}^r$ is amenable, hence $\Gamma_{[g]}$ contains a normal infinite amenable subgroup. The statement of the lemma now follows from Theorem 4.4. □

4.3 Proof of Theorem 2.16

Fix $[(\Sigma, f)] \in \mathsf{Surfaces}(w_1, \ldots, w_\ell)$ and let $\Gamma = \mathrm{MCG}(f)$. Recall that Theorem 2.16 states that $\chi^{(2)}(\Gamma)$ is well-defined and is given by a finite alternating sum over the set $\overline{\mathrm{MATCH}}^*(w_1, \ldots, w_\ell; \Sigma, f)$ of matchings of the letters of $w_1, \ldots, w_\ell$. There is a natural map from elements of $\mathcal{T} \setminus \mathcal{T}_\infty$ to $\overline{\mathrm{MATCH}}^*(w_1, \ldots, w_\ell; \Sigma, f)$:

Definition 4.9. Define a map $\widetilde{\mathbf{match}}$

$$\widetilde{\mathbf{match}} \colon \mathcal{T} \setminus \mathcal{T}_\infty \to \overline{\mathrm{MATCH}}^*(w_1, \ldots, w_\ell)$$

as follows. The (x, j)-arcs of $[g] \in \mathcal{T} \setminus \mathcal{T}_\infty$ define a matching $\sigma_{x,j}$ between the instances of x^{+1} in $w_1, \ldots, w_\ell$ and the instances of x^{-1}. Define $\widetilde{\mathbf{match}}([g])$ to be the element $\sigma \in \overline{\mathrm{MATCH}}^{\kappa(g)}(w_1, \ldots, w_\ell)$ consisting of the matchings $\{\sigma_{x,j}\}_{x \in B, j \in [\kappa_x]}$.

We remark that if $\sigma = \widetilde{\mathbf{match}}([g])$ then indeed $\sigma_{x,j} \neq \sigma_{x,j+1}$ for $x \in B, j < \kappa_x$: this is guaranteed by **Restriction 3** and the fact that the arcs of g cut Σ into discs.

Lemma 4.10. *The map* $\widetilde{\mathbf{match}}$ *descends to a bijection*

$$\mathbf{match} \colon \quad \mathrm{MCG}(f) \backslash (\mathcal{T} \setminus \mathcal{T}_\infty) \xrightarrow{\cong} \overline{\mathrm{MATCH}}^*(w_1, \ldots, w_\ell; \Sigma, f). \tag{4.3}$$

Proof. It is obvious that $\widetilde{\mathbf{match}}$ is invariant under the action of $\mathrm{MCG}(f)$, hence $\mathbf{match}$ is well defined. Since every element $[g] \in \mathcal{T} \setminus \mathcal{T}_\infty$ fills Σ (its arcs cut Σ into discs), it is clear that $(\Sigma, f) \sim (\Sigma_\sigma, f_\sigma)$ where $\sigma = \widetilde{\mathbf{match}}([g])$, using a homeomorphism $\rho \colon \Sigma \to \Sigma_\sigma$ taking g to the transverse map f_σ (so $g \circ \rho^{-1}$ and f_σ are isotopic as transverse maps - recall Definitions 2.7 and 2.11). This shows (i) that $\widetilde{\mathbf{match}}([g])$ indeed belongs to $\overline{\mathrm{MATCH}}^*(w_1, \ldots, w_\ell; \Sigma, f)$ (and not only to $\overline{\mathrm{MATCH}}^*(w_1, \ldots, w_\ell)$), and (ii) that if $\widetilde{\mathbf{match}}([g_1]) = \widetilde{\mathbf{match}}([g_2])$ then $[g_1]$ and $[g_2]$ are in the same $\mathrm{MCG}(f)$-orbit of $\mathcal{T} \setminus \mathcal{T}_\infty$, hence (4.3) is injective.

Finally, to see (4.3) is surjective, notice that for every $\sigma \in \overline{\mathrm{MATCH}}^*(w_1, \ldots, w_\ell; \Sigma, f)$, if ρ is the homeomorphism showing the equivalence of $(\Sigma, f) \sim (\Sigma_\sigma, f_\sigma)$ as above, then $[f_\sigma \circ \rho] \in \mathcal{T} \setminus \mathcal{T}_\infty$, and its image through $\widetilde{\mathbf{match}}$ is σ. $\square$

It follows from Claim 2.10 that $\overline{\mathrm{MATCH}}^*(w_1, \ldots, w_\ell; \Sigma, f)$ is finite, hence:

Corollary 4.11. *There are finitely many* $\mathrm{MCG}(f)$*-orbits in* $\mathcal{T} \setminus \mathcal{T}_\infty$.

We can now prove Theorem 2.16.

Proof of Theorem 2.16. The polysimplicial complex $|\mathcal{T}|_{\mathrm{poly}}$ is a Γ-CW-complex for $\Gamma = \mathrm{MCG}(f)$ by Lemma 4.5. The isotropy groups of Γ in its action on $|\mathcal{T}|_{\mathrm{poly}}$ are either trivial if $[g] \in \mathcal{T} \setminus \mathcal{T}_\infty$ (Lemma 4.7) or infinite if $[g] \in \mathcal{T}_\infty$ (Lemma 4.8). Since $\Gamma \backslash (\mathcal{T} \setminus \mathcal{T}_\infty)$ is finite (Corollary 4.11), we have that

$$m\left(|\mathcal{T}|_{\mathrm{poly}}, \Gamma\right) = \sum_{[g] \in \Gamma \backslash \mathcal{T}} \left|\Gamma_{[g]}\right|^{-1} = \sum_{[g] \in \Gamma \backslash (\mathcal{T} \setminus \mathcal{T}_\infty)} \left|\Gamma_{[g]}\right|^{-1}$$

is finite. From Theorem 4.2 we deduce that $\chi^{(2)}\left(|\mathcal{T}|_{\mathrm{poly}}, \Gamma\right)$ is well defined and given by

$$\begin{aligned} \chi^{(2)}\left(|\mathcal{T}|_{\mathrm{poly}}, \Gamma\right) &= \sum_{[g] \in \Gamma \backslash \mathcal{T}} (-1)^{\dim(\mathrm{polysim}[g])} \left|\Gamma_{[g]}\right|^{-1} = \sum_{[g] \in \Gamma \backslash (\mathcal{T} \setminus \mathcal{T}_\infty)} (-1)^{|\kappa(g)|} \left|\Gamma_{[g]}\right|^{-1} \\ &= \sum_{[g] \in \Gamma \backslash (\mathcal{T} \setminus \mathcal{T}_\infty)} (-1)^{|\kappa(g)|} = \sum_{\sigma \in \overline{\mathrm{MATCH}}^*(w_1, \ldots, w_\ell; \Sigma, f)} (-1)^{|\kappa(\sigma)|}, \end{aligned}$$

where the last equality follows from Lemma 4.10, as the bijection maps the orbit of $[g] \in \mathcal{T} \setminus \mathcal{T}_\infty$ to a set of matchings σ with $\kappa(\sigma) = \kappa(g)$.

Finally, Theorem 3.14 and Lemmas 4.7 and 4.8 show that the assumptions of Lemma 4.3 hold for the action of $\Gamma = \mathrm{MCG}(f)$ on $|\mathcal{T}|_{\mathrm{poly}}$. As $m\left(|\mathcal{T}|_{\mathrm{poly}}, \Gamma\right)$ is finite, we conclude that

$$\chi^{(2)}(\Gamma) = \chi^{(2)}\left(|\mathcal{T}|_{\mathrm{poly}}, \Gamma\right) = \sum_{\sigma \in \overline{\mathrm{MATCH}}^*(w_1,\ldots,w_\ell;\Sigma,f)} (-1)^{|\kappa(\sigma)|}.$$

□

This completes the proof of Theorem 2.16, and hence of our main Theorem 1.7 and of Theorem 1.4.

4.4 Incompressible maps and the proof of Theorem 1.6

Definition 4.12 ([Bro82, Page 247])**.** If G is a discrete group and X is a G-CW-complex such that G acts freely on X, X is contractible, and $G\backslash X$ is a finite CW-complex, then one defines the *Euler characteristic of* G to be

$$\chi(G) \stackrel{\mathrm{def}}{=} \chi(G\backslash X)$$

where the right hand side is the topological Euler characteristic. Since $G\backslash X$ is a $K(G,1)$-space for G, hence unique up to weak homotopy equivalence, this definition does not depend on X.

Recall from Definition 1.5 that $[(\Sigma, f)] \in \mathsf{Surfaces}(w_1, \ldots, w_\ell)$ is called incompressible if it admits no null-curves.

Lemma 4.13. *$[(\Sigma, f)]$ is incompressible if and only if $\mathcal{T}_\infty(\Sigma, f)$ is empty.*

Proof. If (Σ, f) admits a null-curve γ, one can start with an arbitrary element $[g]$ of $\mathcal{T}$ and surger g using H-moves to remove its intersections with γ, similarly to the proof of Proposition 3.20 with γ playing the role of the null-arc. It is easy to check the resulting $[g']$ is in $\mathcal{T}_\infty$. In the other direction, the arcs and curves of any element $[g] \in \mathcal{T}_\infty$ are disjoint from some essential simple closed curve, which is thus a null-curve of (Σ, f). □

Recall that Theorem 1.6 says that an incompressible (Σ, f) admits a finite complex as a $K(\Gamma, 1)$-space for $\Gamma = \mathrm{MCG}(f)$, and that $\chi(\Gamma) = \chi^{(2)}(\Gamma)$.

Proof of Theorem 1.6. By Lemmas 4.5, 4.7 and 4.13, Γ acts freely on the Γ-CW-complex $|\mathcal{T}|_{\mathrm{poly}}$, and by Corollary 4.11 the quotient $\Gamma\backslash|\mathcal{T}|_{\mathrm{poly}}$ is finite. As $|\mathcal{T}|_{\mathrm{poly}}$ is contractible (Theorem 3.14) we obtain that $\Gamma\backslash|\mathcal{T}|_{\mathrm{poly}}$ is the sought-after $K(\Gamma,1)$ -complex. Hence $\chi(\Gamma) = \chi\left(\Gamma\backslash|\mathcal{T}|_{\mathrm{poly}}\right)$ is well defined. Moreover, the proof of the Theorem 2.16 in Section 4.3 shows that $\chi(\Gamma) = \chi^{(2)}(\Gamma)$. □

Remark 4.14. Note that the $K(\Gamma,1)$-complex we obtained as a quotient in the last proof can also be constructed directly as a cell complex with a cell for every $\sigma \in \overline{\mathrm{MATCH}}^*(w_1, \ldots, w_\ell, \Sigma, f)$, in an analogous way to Definition 3.9 of the complex of transverse maps. The example of the single incompressible map for $w = [x,y][x,z]$, where there are two vertices connected by two parallel edges, illustrates that this is not always a polysimplicial complex. However, the set $\overline{\mathrm{MATCH}}^*(w_1, \ldots, w_\ell, \Sigma, f)$ also has a natural partial order defined by forgetting proper subsets of the matchings for every $x \in B$. We can thus realize this $K(\Gamma,1)$ also as the order complex $\left|\overline{\mathrm{MATCH}}^*(w_1, \ldots, w_\ell, \Sigma, f)\right|$, which is a genuine simplicial complex.

We end this subsection with a bound on the dimension of the $K(\Gamma,1)$-complex we constructed:

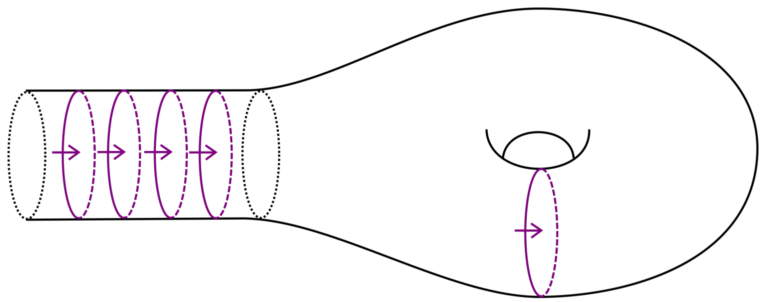

Figure 4.1: This figure shows part of a transverse map, the surface extends to the left where there may be other arcs and curves making up the map. Let $x \in B$. The transverse map shown has $\kappa_x = 0$ and purple curves are $(x, 0)$-curves.

Corollary 4.15. *If $w_1, \ldots, w_\ell$ are all cyclically reduced and different than 1, then the $K(\Gamma, 1)$-space we constructed has dimension at most $-\chi(\Sigma)$.*

Proof. The $K(\Gamma, 1)$-space is a quotient of $\mathcal{T}(\Sigma, f)$, and therefore has the same dimension, which is bounded by $-\chi(\Sigma)$ – see Lemma 3.12 and Remark 3.13. □

4.5 Non-finiteness of $\mathrm{MCG}(f)\backslash\mathcal{T}$: why (the proof of) Theorem 1.6 fails for compressible maps

When (Σ, f) is compressible, the subposet $\mathcal{T}_\infty$ is non-empty (Lemma 4.13), hence the action of $\Gamma = \mathrm{MCG}(f)$ on $|\mathcal{T}|_{\mathrm{poly}}$ is not free, and the quotient is not a $K(\Gamma, 1)$. Still, the ordinary Euler characteristic of a group is defined in much more general cases then the one based on a finite $K(\Gamma, 1)$-space as in Definition 4.12 – see [Bro82, Chapter IX]. For example, one could hope to use the following:

Theorem ([Bro82, Proposition IX.7.3(e')])**.** *Let G be a discrete group, and let X be a contractible G-CW-complex such that $G\backslash X$ has finitely many cells and such that the isotropy group G_c of every cell c "has finite homological type" (see [Bro82, Page 246]). Then $\chi(G)$ is defined and satisfies*

$$\chi(G) = \sum_{[c]\in G\backslash X} (-1)^{\dim c} \chi(G_c).$$

It is not too difficult to show that when $G = \Gamma = \mathrm{MCG}(f)$ and $X = |\mathcal{T}|_{\mathrm{poly}}$, all the assumptions in this theorem hold, except for the assumption that $\Gamma\backslash|\mathcal{T}|_{\mathrm{poly}}$ has finitely many cells. It turns out, perhaps counter-intuitively, that indeed this latter assumption often fails:

Abelian neck phenomenon

Let g be the transverse map which is partially depicted in Figure 4.1, and (Σ, f) be such that $[g] \in \mathcal{T}(\Sigma, f)$. The key feature of g is that there is a null-curve (e.g., one of the dotted black lines) that *separate*s a subsurface that is mapped by f at the level of π_1 to the cyclic group $\langle x \rangle$. In terms of our picture, this can be seen as the curves that appear in this subsurface are associated to only one generator x. Note also in our picture we have illustrated a 'neck' region bounded by two black dotted curves that contains 4 parallel and codirected $(x, 0)$-curves. Assume for the sake of clarity that any curve that could be drawn in the neck region is indeed drawn there. One could modify this transverse map by changing the number of the repeated curves in the neck region.

Lemma 4.16. *No matter how many parallel codirected $(x, 0)$-curves are placed in the neck region of g, the resulting transverse map still realizes (Σ, f) and thus represents an element of $\mathcal{T}(\Sigma, f)$.*

Proof. Call the rightmost $(x, 0)$-arc in the neck α and the $(x, 0)$-curve in the right part of Figure 4.1 β. Consider an H-move along a piece of arc connecting α to β (and arrives to both from their positive side). This results in a strict transverse map in $\mathcal{T}(\Sigma, f)$ which is the same as g except that α is omitted. This shows that alternating the number of $(x, 0)$-curve in the neck in Figure 4.1 does not take us out of $\mathcal{T}(\Sigma, f)$. $\square$

Corollary 4.17. *With $\mathcal{T} = \mathcal{T}(\Sigma, f)$ as in Lemma 4.16, obtained from Figure 4.1,* $\mathrm{MCG}(f)\backslash\mathcal{T}$ *is not finite.*

Proof. We can create elements in $\mathcal{T}$ with an unbounded number of curves, and the number of curves is a $\mathrm{MCG}(f)$-invariant. $\square$

This is not the most general version of this phenomenon: the neck could for example be replaced by a collection of disjoint annuli that cut from Σ some subsurface with π_1 mapped by f to a non-trivial cyclic subgroup of $\mathbf{F}_r$. Our aim here is to give an illustrative example.

5 Further applications and consequences

We specify here three interesting applications of our results and techniques, regarding the stable commutator length of a word, the complete classification of all incompressible solutions in $\mathsf{Surfaces}(w_1, \ldots, w_\ell)$, and the cohomological dimension of $\mathrm{MCG}(f)$. Let us also mention that our construction of a finite $K(\Gamma, 1)$-space for $\Gamma = \mathrm{MCG}(f)$ when $[(\Sigma, f)]$ is incompressible also enables one to write explicit finite presentations for Γ: consult [MP15, Pages 57-59].

5.1 Stable commutator length

Recall that Corollary 1.11 states that the w-measures on $\{\mathcal{U}(n)\}_{n\in\mathbb{N}}$ determine $\mathrm{scl}(w)$, the stable commutator length of $w \in \mathbf{F}_r$, defined in (1.7). In this subsection we explain how this result follows from Theorem 1.7 and from Calegari's rationality theorem.

Calegari's theorem, which is the main result of [Cal09b], says that $\mathrm{scl}(w)$ is rational for every $w \in [\mathbf{F}_r, \mathbf{F}_r]$. First, it is shown that $\mathrm{scl}(\mathrm{w})$ is equal to the infimum of $\frac{-\chi(\Sigma)}{2|j_1+\ldots+j_\ell|}$ over all possible $j_1, \ldots, j_\ell \in \mathbb{Z}$ and (Σ, f) admissible for $w^{j_1}, \ldots, w^{j_\ell}$ [Cal09b, Lemma 2.6]. The proof goes through showing the existence of "extremal surfaces" for w: a surface attaining the infimum. Moreover, by [Cal09b, Lemma 2.7], this extremal surface can be taken to be admissible for $w^{j_1}, \ldots, w^{j_\ell}$ with $j_1, \ldots, j_\ell > 0$. By definition of extremal surface, Σ has maximal Euler characteristic for $w^{j_1}, \ldots, w^{j_\ell}$, namely, $\chi(\Sigma) = \chi_{\max}\left(w^{j_1}, \ldots, w^{j_\ell}\right)$. In fact, every surface which is admissible for $w^{j_1}, \ldots, w^{j_\ell}$ with Euler characteristic $\chi_{\max}\left(w^{j_1}, \ldots, w^{j_\ell}\right)$ is extremal. By [Cal09b, Lemma 2.9], the maps associated with extremal surfaces are π_1-injective, namely, if $\gamma \subset \Sigma$ is a non-nullhomotopic closed curve, then $f(\gamma)$ is not nullhomotopic. Note that this condition is stronger than incompressibility, which only deals with *simple* closed curves. The crux of the matter is the following lemma:

Lemma 5.1. *If (Σ, f) is π_1-injective, then* $\mathrm{MCG}(f)$ *is trivial.*

Proof. The outline of the argument here is that if $[\rho] \in \mathrm{MCG}(f)$ then $[\rho]_* \in \mathrm{Aut}(\pi_1(\Sigma))$ fixes f_*, and since f_* is injective, this means that $[\rho]_*$ must be the identity. By a variation of the Dehn-Nielsen-Baer Theorem, it follows that $[\rho]$ is the identity.

In more detail, assume that Σ is connected (the general cases easily follows). Recall that $v_1 \in \partial_1\Sigma$ is one of the ℓ marked points at $\partial\Sigma$, and let $G = \pi_1(\Sigma, v_1)$. If $\ell = 1$, the Dehn-Nielsen-Baer Theorem (see Page 7 and [MP15, Thm 2.4]) yields what we need. If $\ell \geq 2$, consider an arc $\gamma \subset \Sigma$ connecting v_1 and v_ℓ. Because $[\rho] \in \mathrm{MCG}(\Sigma)$ fixes the marked points, we must have that $\rho(\gamma)$ is homotopic relative to $\{v_1, v_\ell\}$ to $\beta * \gamma$, where β is a closed, not necessarily simple, curve based at v_1 and "$*$" stands for concatenation. Inside $\mathbf{F}_r$ we have

$$f_*[\gamma] = f_*[\rho(\gamma)] = f_*[\beta * \gamma] = f_*[\beta] \cdot f_*[\gamma]$$

hence $f_*[\beta] = 1$ which means that β is nullhomotopic by π_1-injectivity. Hence we can assume without loss of generality that ρ fixes γ, and we can analyze ρ on Σ', the surface obtained from Σ by cutting along γ. Since Σ' has only $\ell - 1$ boundary components, we are done by induction. $\square$

Proof of Corollary 1.11. By Lemma 5.1 and the discussion preceding it, if one of the extremal surfaces of w is admissible for $w^{j_1}, \ldots, w^{j_\ell}$ with $j_1, \ldots, j_\ell > 0$, then Theorem 1.7 translates in this case to

$$\mathcal{T}r_{w^{j_1},\ldots,w^{j_\ell}}(n) = n^{\chi_{\max}\left(w^{j_1},\ldots,w^{j_\ell}\right)} \cdot K + O\left(n^{\chi_{\max}\left(w^{j_1},\ldots,w^{j_\ell}\right)-2}\right), \tag{5.1}$$

where K is the number of highest-Euler-characteristic surfaces in $\mathsf{Surfaces}\left(w^{j_1}, \ldots, w^{j_\ell}\right)$. Note that (5.1) is strictly positive for large enough n. Hence,

$$\frac{-\lim_{n\to\infty}\log_n\left|\mathcal{T}r_{w^{j_1},\ldots,w^{j_\ell}}(n)\right|}{2(j_1+\ldots+j_\ell)} = \frac{-\chi_{\max}\left(w^{j_1},\ldots,w^{j_\ell}\right)}{2(j_1+\ldots+j_\ell)} = \mathrm{scl}(w).$$

On the other hand, for an arbitrary $\ell > 0$ and $j_1, \ldots, j_\ell > 0$ we have

$$\frac{-\lim_{n\to\infty}\log_n\left|\mathcal{T}r_{w^{j_1},\ldots,w^{j_\ell}}(n)\right|}{2(j_1+\ldots+j_\ell)} \geq \frac{-\chi_{\max}\left(w^{j_1},\ldots,w^{j_\ell}\right)}{2(j_1+\ldots+j_\ell)} \geq \mathrm{scl}(w).$$

This proves (1.8). $\square$

Corollary 5.2. *If $\mathrm{scl}(w_1) \neq \mathrm{scl}(w_2)$ then for every large enough n, the w_1-measure on $\mathcal{U}(n)$ is different from the w_2-measure on $\mathcal{U}(n)$. In particular, if $w_1 \in [\mathbf{F}_r, \mathbf{F}_r]$ and $w_2 \notin [\mathbf{F}_r, \mathbf{F}_r]$ then they induce different measures on $\mathcal{U}(n)$ for almost all n.*

Proof. Assume without loss of generality that $\mathrm{scl}(w_1) < \mathrm{scl}(w_2)$, and let $j_1, \ldots, j_\ell > 0$ be so that $w_1^{j_1}, \ldots, w_1^{j_\ell}$ admit an extremal surface. Then by the above discussion, $\mathcal{T}r_{w_1^{j_1},\ldots,w_1^{j_\ell}}(n)$ is strictly larger than $\mathcal{T}r_{w_2^{j_1},\ldots,w_2^{j_\ell}}(n)$ for any large enough n. In particular, if w_2 is not balanced, i.e. $w_2 \notin [\mathbf{F}_r, \mathbf{F}_r]$ and $\mathrm{scl}(w_2) = \infty$, then nor is the set $w_2^{j_1}, \ldots, w_2^{j_\ell}$ balanced as we assume $j_1, \ldots, j_\ell > 0$. By Claim 2.1, $\mathcal{T}r_{w_2^{j_1},\ldots,w_2^{j_\ell}}(n) \equiv 0$ for every n. $\square$

5.2 Classifying all incompressible solutions to generalized commutator equation

Since the late 1970's there are known algorithms to determine the commutator length of a given word $w \in [\mathbf{F}_r, \mathbf{F}_r]$ [Edm75, GT79, Cul81] and also to find at least one representative from every equivalence class of solutions to $[u_1, v_1] \cdots [u_g, v_g] = w$ with $g = \mathrm{cl}(w)$ [Cul81, Section 4.2]. In fact, the algorithm in [Cul81] uses matchings of letters of w as in Proposition 2.13. Our analysis and techniques expand Culler's algorithm to yield a clear description of the set of classes of solutions and, in particular, a direct way to distinguish them from each other.

Consider the poset $P = \overline{\mathrm{MATCH}}^{|\kappa|\leq 1}(w_1, \ldots, w_\ell)$ consisting of sets of matchings for $w_1, \ldots, w_\ell$ as in Section 2, where $|\kappa| \overset{\mathrm{def}}{=} \sum_{x\in B}\kappa_x \leq 1$ and $\sigma_{x,0} \neq \sigma_{x,1}$ whenever $\kappa_x = 1$, and with partial order $\sigma_0 \prec \sigma_1$ whenever $|\kappa(\sigma_0)| = 0$, $|\kappa(\sigma_1)| = 1$ and σ_0 is obtained from σ_1 by deleting one of the two x-matchings for the $x \in B$ with $\kappa_x(\sigma_1) = 1$. Recall the definition of $\chi(\sigma)$ from Definition 2.7. Construct a graph $G(w_1, \ldots, w_\ell)$ with vertices the elements of P and an edge (σ_0, σ_1) whenever $\sigma_0 \prec \sigma_1$ *and* $\chi(\sigma_1) = \chi(\sigma_2)$. We say a component C of $G(w_1, \ldots, w_\ell)$ is *downward-closed* if every vertex σ_1 of C with $\kappa(\sigma_1) = 1$ has two neighbors: the two elements of P that are strictly smaller. Recall the notation Σ_σ and f_σ from Definitions 2.7 and 2.11.

Proposition 5.3. *The map* $\varphi\colon P = \overline{\mathrm{MATCH}}^{|\kappa|\leq 1} \to \mathsf{Surfaces}\,(w_1,\dots,w_\ell)$ *given by*

$$\sigma \mapsto [(\Sigma_\sigma, f_\sigma)]$$

induces a bijection between the downward-closed components of $G\,(w_1,\dots,w_\ell)$ *and the incompressible pairs in* $\mathsf{Surfaces}\,(w_1,\dots,w_\ell)$.

Proof. First, φ is constant on connected components of $G\,(w_1,\dots,w_\ell)$: indeed, assume that $\sigma_0 \prec \sigma_1$ *with* $\chi\,(\sigma_0) = \chi\,(\sigma_1)$ and, say, σ_0 is obtained from σ_1 by forgetting the matching $(\sigma_1)_{x,1}$. Then the condition $\chi\,(\sigma_0) = \chi\,(\sigma_1)$ shows forgetting the $(x,1)$ transversion point of the transverse map f_{σ_1} results in a transverse map which is still filling, and thus equal to f_{σ_0}. Hence we can define $\hat{\varphi}$ to be a map from the downward-closed components of $G\,(w_1,\dots,w_\ell)$ to $\mathsf{Surfaces}\,(w_1,\dots,w_\ell)$.

Second, the image of $\hat{\varphi}$ consists of incompressible elements. To see this, let C be a downward-closed component of P. Let $\sigma \in C$ have $|\kappa\,(\sigma)| = 0$. Assume to the contrary that $\varphi\,(\sigma)$ is compressible. Then it admits a null-curve γ which is not disjoint from the matching-edges in Σ_σ (recall that the matching-edges cut Σ_σ to discs). One can start performing H-moves along this null-curve. In an H-move between two x-matching-edges e_1 and e_2 along a piece of γ, one first creates a transverse map g_1 with $|\kappa\,(g_1)| = 1$ (with $f_\sigma \prec g_1$ in $\mathcal{T} = \mathcal{T}\,(\Sigma_\sigma, f_\sigma)$) and then obtains $g_0 \prec_{\mathcal{T}} g_1$ with $|\kappa\,(g_0)| = 0$ which has two fewer intersection points with γ. Because the matching-edges cut Σ_σ to disks, e_1 and e_2 must be distinct, and thus g_1 has only arcs and $[g_1] = [f_{\sigma_1}]$ for some $\sigma_1 \in C$. As C is downward-closed, there is some $\sigma_0 \in C$ with $[g_0] = [f_{\sigma_0}]$. We can continue in the same manner until γ intersects no matching-edges, which is a contradiction.

Third, $\hat{\varphi}$ is the sought-after bijection. Indeed, every incompressible $[(\Sigma, f)] \in \mathsf{Surfaces}\,(w_1,\dots,w_\ell)$ is the φ-image of some component of $G\,(w_1,\dots,w_\ell)$ by Proposition 2.13. If $\sigma \in P$ satisfies $\varphi\,(\sigma) = [(\Sigma, f)]$ and C is the connected component of σ then C is a component of the 1-skeleton of the $K\,(\Gamma,1)$-complex we constructed in the proof of Theorem 1.6 in Section 4.4. In particular, this complex is connected (because its universal cover $|\mathcal{T}|_{\mathrm{poly}}$ is connected), hence so its 1-skeleton is connected. This shows that C is the only component mapping to $[(\Sigma_\sigma, f_\sigma)]$ and that it is downward-closed. □

Alternatively, one could use here a direct argument imitating some ingredients from the proof of Theorem 3.14, as follows. For $[(\Sigma, f)]$ incompressible, show that $\varphi^{-1}\,([(\Sigma, f)])$ is a downward-closed connected component of $G\,(w_1,\dots,w_\ell)$, by taking a maximal system of null-arcs, showing there is a single $\sigma_0 \in \varphi^{-1}\,([(\Sigma, f)])$ with matching edges disjoint from these null-arcs, and showing every other element in the preimage can be connected to σ_0 by H-moves that never leave the same connected component of $G\,(w_1,\dots,w_\ell)$.

5.3 Finiteness of the cohomological dimension of the stabilizer $\mathrm{MCG}\,(f)$

Recall that the cohomological dimension, $\mathrm{cd}(\Gamma)$, of a torsion-free group Γ is the minimal length of a projective resolution of $\mathbf{Z}$ over $\mathbf{Z}\Gamma$ if one exists, and ∞ otherwise. If a group Γ is virtually torsion-free then the virtual cohomological dimension, $\mathrm{vcd}(\Gamma)$, is defined to be $\mathrm{cd}(\Gamma')$ where Γ' is a finite index torsion-free subgroup of Γ; it is a theorem of Serre [Ser71] that the resulting dimension does not depend on the chosen finite index subgroup. As the the following result is not needed for the main results of this paper, we only sketch its proof.

Proposition 5.4. *Let* Σ *be a compact orientable surface with no closed connected components and let* $f\colon \Sigma \to \bigvee^r S^1$ *be a map. Then* $\mathrm{cd}(\mathrm{MCG}(f)) < \infty$.

Sketch of proof. Let $\mathcal{T} = \mathcal{T}(\Sigma, f)$ and $\Gamma = \mathrm{MCG}(f)$. Note that Γ is torsion-free since Σ has no closed components. We use a result that is attributed to Quillen by Serre in [Ser71, Prop. 11(a)]. As $|\mathcal{T}|_{\mathrm{poly}}$ is contractible (Theorem 3.14), Quillen's result says that

$$\mathrm{cd}(\Gamma) \leq \sup_{[g] \in \Gamma \backslash \mathcal{T}} \left(\dim(\mathrm{polysim}\left([g]\right)) + \mathrm{cd}(\mathrm{Stab}_\Gamma(g)\right).$$

Therefore, as $|\mathcal{T}|_{\mathrm{poly}}$ is finite dimensional (Lemma 3.12), it suffices to prove there is an upper bound depending only on the pair (Σ, f) for $\mathrm{cd}(\mathrm{Stab}_\Gamma(g))$ given an arbitrary element g in $\mathcal{T}$.

We now give a quick analysis of these stabilizers. Fix a transverse map g with $[g] \in \mathcal{T}$. Let $\{\Sigma_i\}_{i \in I}$ denote the zones of g which are not annuli bounded by two curves of g. By Euler characteristic argument, I is finite and bounded independently of g. Form Σ_i^* by contracting each end of Σ_i bounded by a curve of g to a point, and mark the new points $W_i \subset \Sigma_i^*$ on their respective surfaces. We denote by $\mathrm{MCG}(\Sigma_i^*, W_i)$ the mapping class group of Σ_i^* that fixes each individual element of W_i.

The subgroup $\Gamma_0 \leq \mathrm{Stab}_\Gamma(g)$ that fixes all the curves in g and their orientations has finite index in $\mathrm{Stab}_\Gamma(g)$, and there is a short exact sequence obtained by restricting mapping classes in Γ_0 to the zones Σ_i:

$$1 \to N \to \Gamma_0 \to H \stackrel{\text{def}}{=} \prod_{i \in I} \mathrm{MCG}(\Sigma_i^*, W_i) \to 1, \tag{5.2}$$

where N is a free abelian group generated by Dehn twists in the curves of g. The reason one obtains the whole of each $\mathrm{MCG}(\Sigma_i^*, W_i)$ as a factor is because g maps each Σ_i to a contractible piece of $\bigvee^r S^1$, and any lift of any element of $\mathrm{MCG}(\Sigma_i^*, W_i)$ to $\mathrm{MCG}(\Sigma)$ can be taken to be the identity outside Σ_i, and therefore preserves the homotopy class of f. Although $\mathrm{MCG}(\Sigma_i^*, W_i)$ could contain torsion, it is virtually torsion-free (see either [Iva02, Theorem 6.8.A] or [Har86, §4]).

Harer proved in [Har86] that for any surface Σ and collection of interior marked points W,

$$\mathrm{vcd}(\mathrm{MCG}(\Sigma, W)) \leq 4g(\Sigma) + 2|\pi_0(\partial\Sigma)| + |W| - 3,$$

where $g(\Sigma)$ is the genus of Σ. Therefore using [Bro82, Prop. VIII.2.4.b] together with an argument as in [Bro82, Proof of Prop. IX.7.3.d] to pass between vcd and cd, one obtains

$$\mathrm{vcd}(H) \leq \sum_{i \in I} \left(4g(\Sigma_i^*) + 2|\pi_0(\partial\Sigma_i^*)| + |W_i| - 3\right) \leq F_1(\Sigma),$$

where $F_1(\Sigma)$ is a bound in terms of Σ which is independent of g. Since H is virtually torsion-free, and Γ_0 has no torsion, we can find torsion-free finite index subgroups Γ_0', H' in Γ_0 and H respectively that form a short exact sequence $1 \to N \to \Gamma_0' \to H' \to 1$. Then Serre's Theorem [Ser71] gives $\mathrm{cd}(\mathrm{Stab}_\Gamma(g)) = \mathrm{cd}(\Gamma_0) = \mathrm{cd}(\Gamma_0')$ and $\mathrm{cd}(H') = \mathrm{vcd}(H)$. We also have $\mathrm{cd}(N) \leq F_2(\Sigma)$ where $F_2(\Sigma)$ is the maximal number of pairwise non-isotopic disjoint simple closed curves on Σ. Now applying [Bro82, Prop. VIII.2.4.b] to the short exact sequence for N, Γ_0', H' we get

$$\mathrm{cd}(\mathrm{Stab}_\Gamma(g)) = \mathrm{cd}(\Gamma_0') \leq \mathrm{cd}(N) + \mathrm{cd}(H') = \mathrm{cd}(N) + \mathrm{vcd}(H) < F_1(\Sigma) + F_2(\Sigma).$$

□

6 Open problems

We mention some open problems that naturally arise from the discussion in this paper.

1. Recall that primitive words are the orbit in $\mathbf{F}_r$ of the single-letter word x under the action of $\mathrm{Aut}(\mathbf{F}_r)$. As mentioned on Page 10, it was shown in [PP15] that only primitive words induce uniform measure on the symmetric group S_n for all n. Is the same true for unitary groups? Namely, if a word induces Haar measure on $\mathcal{U}(n)$ for all n, is the word necessarily primitive? In fact, the following question raised by Tsachik Gelander a few years ago (by private communication) is still open: if a word induces Haar measure on $\mathcal{U}(2)$, is the word necessarily primitive?

2. Fix $j_1, \ldots, j_\ell \in \mathbb{Z}$. Given $w \in \mathbf{F}_r$, is there a nice criterion for determining whether the rational expression for $\mathcal{T}r_{w^{j_1},\ldots,w^{j_\ell}}(n)$ has the same value as for the primitive case when $w = x$? An illustrating example is $\mathcal{T}r_w(n)$ – we know it vanishes outside $[\mathbf{F}_r, \mathbf{F}_r]$, but it is not clear when it vanishes inside $[\mathbf{F}_r, \mathbf{F}_r]$. Another illustrating example is $\mathcal{T}r_{w,w^{-1}}(n)$: when does it differ from 1? Some examples for each are elaborated in Table 1 and on Page 8.

3. Let Σ be a connected, orientable surface with boundary, and let $f \colon \Sigma \to \bigvee^r S^1$. We showed here that $\mathrm{MCG}(f)$ has a well-defined L^2-Euler-characteristic (Theorem 1.4) and a finite cohomological dimension (Proposition 5.4). Does $\mathrm{MCG}(f)$ always have "finite homological type" as defined in [Bro82, Page 246]? And if so, does its ordinary Euler characteristic coincide with the L^2-one?

4. We deduced the rationality of $\mathcal{T}r_{w_1,\ldots,w_\ell}(n)$ in Theorem 2.8 directly from Weingarten calculus. The rationality means that the different L^2-Euler characteristics appearing in Theorem 1.7 "know" about each other. Is it possible to deduce the rationality of $\mathcal{T}r_{w_1,\ldots,w_\ell}(n)$ (i.e., Proposition 1.1) from our main theorem, Theorem 1.7?

5. What can one say about the distribution of $\mathcal{T}r_w(n)$ when w is a long random word in $[\mathbf{F}_r, \mathbf{F}_r]$? For example, what is the distribution of the commutator length of w? Is it true that for most words of a fixed length in $[\mathbf{F}_r, \mathbf{F}_r]$, the stabilizers $\mathrm{MCG}(f)$ of incompressible solutions are trivial?

6. What can one systematically say about the L^2-Euler characteristic of $\mathrm{MCG}(f)$? For which f are they zero, negative, or positive? The case when f is incompressible is a natural starting point. A sufficiently good understanding of this question would allow one to make progress on Conjecture 1.10. The Euler characteristic of the mapping class group of a closed surface was calculated by Harer-Zagier [HZ86] and the *sign* of the Euler characteristic of the mapping class group was re-obtained by McMullen [McM00] by different methods.

References

[AB83] M. F. Atiyah and R. Bott, *The Yang-Mills Equations over Riemann Surfaces*, Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences **308** (1983), no. 1505, 523–615. 11

[AV11] A. Amit and U. Vishne, *Characters and solutions to equations in finite groups*, J. Algebra Appl. **10** (2011), no. 4, 675–686. MR 2834108 10

[Bro82] K. S. Brown, *Cohomology of groups*, Graduate Texts in Mathematics, vol. 87, Springer-Verlag, New York-Berlin, 1982. MR 672956 (83k:20002) 6, 44, 45, 49, 50

[Cal08] D. Calegari, *What is... stable commutator length?*, Notices Amer. Math. Soc. **55** (2008), no. 9, 1100–1101. MR 2451345 10

[Cal09a] ______, *scl*, MSJ Memoirs, vol. 20, Mathematical Society of Japan, Tokyo, 2009. 7, 10

[Cal09b] ______, *Stable commutator length is rational in free groups*, Journal of the American Mathematical Society **22** (2009), no. 4, 941–961. MR 2525776 (2010k:57002) 10, 46

[CG86] J. Cheeger and M. Gromov, *L_2-cohomology and group cohomology*, Topology **25** (1986), no. 2, 189–215. MR 837621 40, 41

[Col03] B. Collins, *Moments and cumulants of polynomial random variables on unitary groups, the Itzykson-Zuber integral, and free probability*, International Mathematics Research Notices **2003** (2003), no. 17, 953–982. 2, 13, 14, 19

[CŚ06] B. Collins and P. Śniady, *Integration with respect to the Haar measure on unitary, orthogonal and symplectic group*, Comm. Math. Phys. **264** (2006), no. 3, 773–795. MR 2217291 (2007c:60009) 2, 13, 14, 15

[Cul81] M. Culler, *Using surfaces to solve equations in free groups*, Topology **20** (1981), no. 2, 133–145. MR 605653 (82c:20052) 7, 8, 22, 47

[DM69] P. Deligne and D. Mumford, *The irreducibility of the space of curves of given genus*, Inst. Hautes Études Sci. Publ. Math. (1969), no. 36, 75–109. MR 0262240 42

[DS94] P. Diaconis and M. Shahshahani, *On the eigenvalues of random matrices*, Journal of Applied Probability (1994), 49–62. 10

[Edm75] C. C. Edmunds, *On the endomorphism problem for free groups*, Communications in Algebra **3** (1975), no. 1, 1–20. 47

[FM12] B. Farb and D. Margalit, *A primer on mapping class groups*, Princeton Mathematical Series, vol. 49, Princeton University Press, Princeton, NJ, 2012. MR 2850125 (2012h:57032) 7, 42

[Ful97] W. Fulton, *Young tableaux: with applications to representation theory and geometry*, London Mathematical Society Student Texts, vol. 35, Cambridge University Press, 1997. 14

[Gol84] W. M. Goldman, *The symplectic nature of fundamental groups of surfaces*, Advances in Mathematics **54** (1984), no. 2, 200 – 225. 11

[Gol97] W. M. Goldman, *Ergodic theory on moduli spaces*, Ann. of Math. (2) **146** (1997), no. 3, 475–507. MR 1491446 11

[GT79] R. Z. Goldstein and E. C. Turner, *Applications of topological graph theory to group theory*, Mathematische Zeitschrift **165** (1979), no. 1, 1–10. 47

[Har81] W. J. Harvey, *Boundary structure of the modular group*, Riemann surfaces and related topics: Proceedings of the 1978 Stony Brook Conference (State Univ. New York, Stony Brook, N.Y., 1978), Ann. of Math. Stud., vol. 97, Princeton Univ. Press, Princeton, N.J., 1981, pp. 245–251. MR 624817 42

[Har85] J. L. Harer, *Stability of the homology of the mapping class groups of orientable surfaces*, Ann. of Math. (2) **121** (1985), no. 2, 215–249. MR 786348 42

[Har86] ———, *The virtual cohomological dimension of the mapping class group of an orientable surface*, Invent. Math. **84** (1986), no. 1, 157–176. MR 830043 42, 49

[HLS15] C. Y. Hui, M. Larsen, and A. Shalev, *The Waring problem for Lie groups and Chevalley groups*, Israel Journal of Mathematics **210** (2015), no. 1, 81–100. 10

[HZ86] J. L. Harer and D. Zagier, *The Euler characteristic of the moduli space of curves*, Invent. Math. **85** (1986), no. 3, 457–485. MR 848681 (87i:32031) 9, 50

[Iva02] N. V. Ivanov, *Mapping class groups*, Handbook of geometric topology, North-Holland, Amsterdam, 2002, pp. 523–633. MR 1886678 49

[Lab13] F. Labourie, *Lectures on representations of surface groups*, Zurich Lectures in Advanced Mathematics, European Mathematical Society (EMS), Zürich, 2013. MR 3155540 11

[Lüc02] W. Lück, *L^2-invariants: theory and applications to geometry and K-theory*, Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge. A Series of Modern Surveys in Mathematics [Results in Mathematics and Related Areas. 3rd Series. A Series of Modern Surveys in Mathematics], vol. 44, Springer-Verlag, Berlin, 2002. MR 1926649 5, 40, 41

[LZ04] S. K. Lando and A. K. Zvonkin, *Graphs on surfaces and their applications*, Encyclopaedia of Mathematical Sciences, vol. 141, Springer-Verlag, Berlin, 2004, With an appendix by Don B. Zagier, Low-Dimensional Topology, II. MR 2036721 (2005b:14068) 9

[McM00] C. T. McMullen, *The moduli space of Riemann surfaces is Kähler hyperbolic.*, Annals of Mathematics. Second Series **151** (2000), no. 1, 327–357 (eng). 50

[Mon08] G. Mondello, *A remark on the homotopical dimension of some moduli spaces of stable Riemann surfaces*, J. Eur. Math. Soc. (JEMS) **10** (2008), no. 1, 231–241. MR 2349902 42

[MP15] M. Magee and D. Puder, *Word measures on unitary groups*, arXiv preprint 1509.07374 v2, 2015. 2, 7, 8, 10, 12, 46

[MP19a] ______, *Matrix group integrals, surfaces, and mapping class groups II: $O(n)$ and $Sp(n)$*, arXiv preprint arXiv:1904.13106, 2019. 12

[MP19b] ______, *Surface words are determined by word measures on groups*, arXiv preprint arXiv:1902.04873, 2019. 10

[MŚS07] J. A. Mingo, P. Śniady, and R. Speicher, *Second order freeness and fluctuations of random matrices. II. Unitary random matrices*, Adv. Math. **209** (2007), no. 1, 212–240. MR 2294222 (2009c:15027) 11

[Nov17] M. Novaes, *Expansion of polynomial lie group integrals in terms of certain maps on surfaces, and factorizations of permutations*, Journal of Physics A: Mathematical and Theoretical **50** (2017), no. 7, 075201. 19

[NS06] A. Nica and R. Speicher, *Lectures on the combinatorics of free probability*, London Mathematical Society Lecture Note Series, vol. 335, Cambridge University Press, Cambridge, 2006. MR 2266879 (2008k:46198) 14

[Pen88] R. C. Penner, *Perturbative series and the moduli space of Riemann surfaces*, J. Differential Geom. **27** (1988), no. 1, 35–53. MR 918455 (89h:32045) 9

[PP15] D. Puder and O. Parzanchevski, *Measure preserving words are primitive*, Journal of the American Mathematical Society **28** (2015), no. 1, 63–97. MR 3264763 2, 10, 49

[Pud14] D. Puder, *Primitive words, free factors and measure preservation*, Israel J. Math. **201** (2014), no. 1, 25–73. MR 3265279 16

[Qui73] D. Quillen, *Higher algebraic K-theory. I*, Algebraic K-theory, I: Higher K-theories (Proc. Conf., Battelle Memorial Inst., Seattle, Wash., 1972), Springer, Berlin, 1973, pp. 85–147. Lecture Notes in Math., Vol. 341. MR 0338129 38

[Răd06] F. Rădulescu, *Combinatorial aspects of Connes's embedding conjecture and asymptotic distribution of traces of products of unitaries*, Proceedings of the Operator Algebra Conference, Bucharest, Theta Foundation, 2006. 11

[Ser71] J. P. Serre, *Cohomologie des groupes discrets*, Prospects in mathematics (Proc. Sympos., Princeton Univ., Princeton, N.J., 1970), Princeton Univ. Press, Princeton, N.J., 1971, pp. 77–169. Ann. of Math. Studies, No. 70. MR 0385006 48, 49

[Sha13] A. Shalev, *Some results and problems in the theory of word maps*, Erdős Centennial (Bolyai Society Mathematical Studies) (L. Lovász, I. Ruzsa, V.T. Sós, and D. Palvolgyi, eds.), Springer, 2013, pp. 611–650. 10

[tD72] T. tom Dieck, *Orbittypen und äquivariante Homologie. I*, Arch. Math. (Basel) **23** (1972), 307–317. MR 0310919 41

[tD87] ———, *Transformation groups*, De Gruyter Studies in Mathematics, vol. 8, Walter de Gruyter & Co., Berlin, 1987. MR 889050 41

[tH74] G. 't Hooft, *A planar diagram theory for strong interactions*, Nuclear Physics B **72** (1974), no. 3, 461 – 473. 9

[VDN92] D. V. Voiculescu, K. J. Dykema, and A. Nica, *Free random variables*, CRM Monograph Series, vol. 1, American Mathematical Society, Providence, RI, 1992, A noncommutative probability approach to free products with applications to random matrices, operator algebras and harmonic analysis on free groups. MR 1217253 (94c:46133) 11

[Voi91] D. Voiculescu, *Limit laws for random matrices and free products*, Invent. Math. **104** (1991), no. 1, 201–220. MR 1094052 (92d:46163) 11

[von29] J. von Neumann, *Zur allgemeinen Theorie des Maßes.*, Fundam. Math. **13** (1929), 73–116 (German). 42

[Wal88] J. W. Walker, *Canonical homeomorphisms of posets*, European Journal of Combinatorics **9** (1988), no. 2, 97–107. 28

[Wei78] D. Weingarten, *Asymptotic behavior of group integrals in the limit of infinite rank*, Journal of Mathematical Physics **19** (1978), no. 5, 999–1001. 2, 13

[Wit91] E. Witten, *On quantum gauge theories in two dimensions*, Comm. Math. Phys. **141** (1991), no. 1, 153–209. MR 1133264 (93i:58164) 11

[Xu97] F. Xu, *A random matrix model from two dimensional Yang-Mills theory*, Communications in mathematical physics **190** (1997), no. 2, 287–307. 2, 13

Michael Magee,
Department of Mathematical Sciences,
Durham University,
Lower Mountjoy, DH1 3LE Durham,
United Kingdom
`michael.r.magee@durham.ac.uk`

Doron Puder,
School of Mathematical Sciences,
Tel Aviv University,
Tel Aviv, 6997801, Israel
`doronpuder@gmail.com`