

LENGTHS OF SADDLE CONNECTIONS ON RANDOM TRANSLATION SURFACES OF LARGE GENUS

HOWARD MASUR, KASRA RAFI, AND ANJA RANDECKER

ABSTRACT. We determine the distribution of the number of saddle connections in a random translation surface of large genus. More specifically, for genus g going to infinity, the number of saddle connections with lengths in a given interval $[a/g, b/g]$ converges in distribution to a Poisson distributed random variable. Furthermore, the numbers of saddle connections associated to disjoint intervals are independent.

1. INTRODUCTION

We study the distribution of short saddle connections on a random large-genus translation surface. This is part of a continuing effort to understand the geometry of a random surface as the genus of the surface goes to infinity. Much is known about the shape of random hyperbolic surfaces but most questions about random translation surfaces remain open. Here, we aim to emulate the celebrated results of Mirzakhani–Petri [MP19] regarding the number of short curves in a random large-genus hyperbolic surface.

The space of translation surfaces of genus g can be identified with the space of abelian differentials (X, ω) where X is a Riemann surface of genus g and ω is a holomorphic 1-form on X . The space of abelian differentials can be decomposed into strata depending on the number and the order of zeros of ω . Let $\mathcal{H}_g = \mathcal{H}_g(1, 1, \dots, 1)$ be the principle stratum of unit-area abelian differentials on a closed surface of genus g where there are $2g - 2$ zeros of order 1. We equip the stratum with the normalized Lebesgue measure $\text{Vol}(\cdot)$ as in [Mas82, Vee82]. Then \mathcal{H}_g has a finite total volume [Mas82, Vee82] and we can define the probability of a measurable subset $E \subseteq \mathcal{H}_g$ by

$$\mathbb{P}_g(E) = \frac{\text{Vol}(E)}{\text{Vol}(\mathcal{H}_g)}.$$

Since the total area of a surface in \mathcal{H}_g is 1 and the number of zeros of an abelian differential in \mathcal{H}_g is $2g - 2$, the expected number of saddle connections of length at most ϵ on a random translation surface in \mathcal{H}_g goes to infinity as $g \rightarrow \infty$. Therefore, we need to choose the right scale at which the expected number of saddle connections is a finite, positive real number. We show that this correct scale is $1/g$ in the following sense. Given $(X, \omega) \in \mathcal{H}_g$ and an interval $[a, b] \subseteq \mathbb{R}_+$, let $N_{g,[a,b]}(X, \omega)$ denote the number of saddle connections on (X, ω) with lengths in the interval $[a/g, b/g]$.

Theorem 1.1. *Let $[a_1, b_1], [a_2, b_2], \dots, [a_k, b_k] \subseteq \mathbb{R}_+$ be disjoint intervals. Then, as $g \rightarrow \infty$, the vector of random variables*

$$(N_{g,[a_1,b_1]}, \dots, N_{g,[a_k,b_k]}) : \mathcal{H}_g \rightarrow \mathbb{N}_0^k$$

converges jointly in distribution to a vector of random variables with Poisson distributions of means $\lambda_{[a_i,b_i]}$, where

$$\lambda_{[a_i,b_i]} = 8\pi(b_i^2 - a_i^2)$$

for $i = 1, \dots, k$. That is,

$$\lim_{g \rightarrow \infty} \mathbb{P}_g (N_{g,[a_1,b_1]} = n_1, \dots, N_{g,[a_k,b_k]} = n_k) = \prod_{i=1}^k \frac{\lambda_{[a_i,b_i]}^{n_i} e^{-\lambda_{[a_i,b_i]}}}{n_i!}.$$

Theorem 1.1 has the following consequence: Let $l_{\min}(X, \omega)$ be the length of the shortest saddle connection on (X, ω) . Then

$$\lim_{g \rightarrow \infty} \mathbb{P}_g (l_{\min} < \epsilon/g) = 1 - \mathbb{P}(N_{g,[0,\epsilon]} = 0) = 1 - e^{-\lambda^{[0,\epsilon]}} \approx 8\pi\epsilon^2 \quad \text{as } \epsilon \rightarrow 0.$$

This is in line with the fact that for a fixed stratum \mathcal{H} and a small $\epsilon > 0$, the thin part $\mathcal{H}_{\text{thin}}^\epsilon \subseteq \mathcal{H}$ which contains all translation surfaces with a saddle connection of length at most ϵ satisfies

$$\text{Vol}(\mathcal{H}_{\text{thin}}^\epsilon) = O(\epsilon^2) \cdot \text{Vol}(\mathcal{H}).$$

For one short saddle connection, this can be proven using Siegel–Veech constants. However, also for the set of translation surfaces with two short saddle connections, there exists a similar result on asymptotics by [MS91] which can be made more concrete with Theorem 1.1.

Other strata. For a given genus, the principal stratum has the highest dimension and includes all other strata for that genus on its boundary. Thus, although Theorem 1.1 is stated for the principal stratum, it also applies to the space of all unit-area translation surfaces of a given genus. For non-principal strata, the result does not hold universally; however, in certain cases, similar conclusions can be drawn, which we now explore.

The first variation is for strata where the number of non-simple zeros is growing slow enough.

Theorem 1.2. *Let $f, \ell: \mathbb{N}_0 \rightarrow \mathbb{R}$ be functions such that $\frac{f(g)^2 \cdot \ell(g)}{g} \rightarrow 0$ for $g \rightarrow \infty$. If \mathcal{H}_g is replaced by the stratum $\mathcal{H}_g(m_1, m_2, \dots, m_{\ell(g)}, 1, \dots, 1)$ of unit-area translation surfaces of genus g with at most $\ell(g)$ zeros which are not simple and $m_1, \dots, m_{\ell(g)} \leq f(g)$, then the same statement as in Theorem 1.1 is true for the same $\lambda_{[a_i,b_i]} = 8\pi(b_i^2 - a_i^2)$.*

We can also consider strata where none of the zeros is simple but all zeros have the same, fixed order.

Theorem 1.3. *Fix $m \in \mathbb{N}$ odd and consider only such $g \in \mathbb{N}$ for which $2g - 2$ is divisible by m . If \mathcal{H}_g is replaced by the stratum $\mathcal{H}_g(m, m, \dots, m)$ of unit-area translation surfaces of genus g where there are $\frac{2g-2}{m}$ zeros of order m , then the same statement as in Theorem 1.1 is true with means*

$$\lambda_{[a_i,b_i]} = \left(\frac{m+1}{m} \right)^2 \cdot 2\pi(b_i^2 - a_i^2)$$

for $i = 1, \dots, k$.

Other slight variations of the theorem can also be derived with the same method. However, our method does not work for the minimal stratum $\mathcal{H}_g(2g-2)$ when there is only one zero. The reason is that for a surface in the minimal stratum, every saddle connection is a closed loop and cannot be collapsed. However, the scale $1/g$ still seems to be correct.

Question 1.4. *Is there an analogue of Theorem 1.1 for the space $\mathcal{H}_g(2g-2)$?*

History and tools. Large-genus asymptotics in the context of surfaces have been studied for some years primarily for hyperbolic surfaces. Mirzakhani started a program in [Mir13] in which she determined the expected systole, Cheeger constant, diameter, and other geometric properties of random surfaces of large genus. Subsequently, Mirzakhani–Petri worked out in [MP19] the distribution of the number of short curves in a random hyperbolic surface of large genus which is the inspiration for this article.

For translation surfaces, less is still known than for hyperbolic surfaces. However, for the volumes of strata of translation surfaces, there has been a recent increase in results. Based on work of Eskin–Okounkov [EO01], Eskin–Zorich predicted in [EZ15] the volumes of strata of translation surfaces and Aggarwal confirmed these [Agg20]. A more precise error term for the volume estimates has been determined by Chen–Möller–Sauvaget–Zagier [CMSZ20]. Also for half-translation surfaces, first results have been obtained by Chen–Möller–Sauvaget [CMS23].

By work of Eskin–Masur–Zorich [EMZ03], volumes of strata are strongly related to Siegel–Veech constants (see Section 4 for definitions and more details). Asymptotics of Siegel–Veech constants for large genus have been calculated by many different authors in recent years, e.g. Chen–Möller–Zagier [CMZ18] for principal strata, Sauvaget [Sau18] for minimal strata, Zorich in an appendix to [Agg20] and Aggarwal [Agg19] for all strata of translation surfaces, and Chen–Möller–Sauvaget–Zagier [CMSZ20] for principal strata of half-translation surfaces.

Furthermore, also geometric properties of random translation surfaces of large genus have been studied: An upper bound for the covering radius is given in [MRR22] and Delecroix–Goujard–Zograf–Zorich have determined geometric and combinatorial properties for large-genus square-tiled surfaces in a series of articles, see [DGZZ23] and the references therein. In an upcoming work, Bowen–Rafi–Vallejos [BRV24] show that a sequence of random translation surfaces with area equal to genus Benjamini–Schramm converges as genus tends to infinity.

For the proof of our main theorem, we will strongly use the recent works on large-genus asymptotics for volumes of strata and for Siegel–Veech constants that are outlined above.

Further questions and remarks. Let $\text{sys}(X, \omega)$ be the length of the shortest closed geodesic in (X, ω) . To examine sys , we need to work at a different scale as the expected number of closed geodesics of length $1/g$ converges to zero. It turns out that the correct scale for this setup is $1/\sqrt{g}$. Let $L_{g,[a,b]}(X, \omega)$ denote the number of closed geodesics on (X, ω) with lengths in the interval $[a/\sqrt{g}, b/\sqrt{g}]$.

Question 1.5. *Does $L_{g,[a,b]}(X, \omega)$ converge in distribution to a random variable with a Poisson distribution?*

We can not directly apply the same methods as in this paper, again because our methods involve the collapsing of saddle connections which does not work for closed geodesics.

Another natural setting is the space of quadratic differentials (or half-translation surfaces). Let $\mathcal{Q}_g = \mathcal{Q}_g(1, 1, \dots, 1)$ be the principle stratum of unit-area quadratic differentials. Again the correct scale seems to be $1/g$.

Question 1.6. *Does an analogue of Theorem 1.1 and/or an analogue of Question 1.5 hold for the space \mathcal{Q}_g ?*

Again, our methods fail for the analogue of Theorem 1.1 since collapsing a saddle connection in a quadratic differential is not a local construction. Furthermore, much less is known about the volumes of strata of quadratic differentials.

Outline of the paper. In the next four sections, we recall the background needed for our arguments: In Section 2, we describe the method of moments which is used to deduce that

a variable is Poisson distributed. In Section 3, we introduce translation surfaces and their strata. The method of Siegel–Veech constants and some relevant values of them are collected in Section 4. In Section 5, we recall the construction for collapsing a saddle connection and describe a variation of it to collapse two saddle connections sharing a zero.

We start the proof of Theorem 1.1 in Section 6 by defining a generic subset of the stratum, on which we describe a simultaneous collapsing procedure in Section 7. This construction is used in Section 8 to determine the factorial moments of $N_{g,[a,b]}$ on the generic subset. In Section 9, we explain how to extend the result from the generic subset to the whole stratum.

We finish with arguments to prove the variations of the main theorem in Section 10.

Acknowledgements. The authors are very grateful to Anton Zorich for many helpful conversations during the work on this project as well as for factors of 2 and detailed comments on an earlier draft of this work.

K. R. was partially supported by NSERC Discovery Grant RGPIN 05507 and the Simons Fellowship Program. The work of A. R. was partially funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 441856315.

2. THE METHOD OF MOMENTS

Similar to the proof in [MP19], we use the method of moments for showing that a random variable is Poisson distributed. Recall that a random variable $N: \Omega \rightarrow \mathbb{N}_0$ on a probability space (Ω, \mathbb{P}) is said to be *Poisson distributed with mean* $\lambda \in (0, \infty)$ if

$$\mathbb{P}(N = k) = \frac{\lambda^k e^{-\lambda}}{k!} \quad \text{for all } k \in \mathbb{N}_0.$$

The moments of a Poisson distributed random variable are slightly complicated terms, so we instead consider the factorial moments here.

Given a random variable $N: \Omega \rightarrow \mathbb{N}_0$ and $r \in \mathbb{N}$, we define the random variable

$$(N)_r = N(N-1)\dots(N-r+1).$$

If its expectation $\mathbb{E}(N)_r$ exists, it is called the r -th *factorial moment* of N .

The r -th factorial moment of a Poisson distributed variable with mean λ is equal to λ^r for all $r \in \mathbb{N}$. It turns out that this is an if and only if condition.

Theorem 2.1 (The method of moments [Bol01, Theorem 1.23]). *Let $\{(\Omega_i, \mathbb{P}_i)\}_{i \in \mathbb{N}}$ be a sequence of probability spaces. For $k \in \mathbb{N}$, let $N_{1,i}, \dots, N_{k,i}: \Omega_i \rightarrow \mathbb{N}_0$ be random variables for all $i \in \mathbb{N}$ and suppose there exist $\lambda_1, \dots, \lambda_k \in (0, \infty)$ such that*

$$\lim_{i \rightarrow \infty} \mathbb{E}((N_{1,i})_{r_1} \dots (N_{k,i})_{r_k}) = \lambda_1^{r_1} \dots \lambda_k^{r_k}$$

for all $r_1, \dots, r_k \in \mathbb{N}$. Then

$$\lim_{i \rightarrow \infty} \mathbb{P}(N_{1,i} = n_1, \dots, N_{k,i} = n_k) = \prod_{j=1}^k \frac{\lambda_j^{n_j} e^{-\lambda_j}}{n_j!}$$

for all $n_1, \dots, n_k \in \mathbb{N}$. In other words, the vector $(N_{1,i}, \dots, N_{k,i}): \Omega_i \rightarrow \mathbb{N}_0^k$ converges jointly in distribution to a vector which is independently Poisson distributed with means $\lambda_1, \dots, \lambda_k$.

3. TRANSLATION SURFACES AND STRATA

A *translation surface* (X, ω) is a pair consisting of a compact, connected Riemann surface X and a holomorphic 1-form ω on X . By integrating the 1-form, we obtain a metric on X which is locally isometric to the Euclidean plane except in a neighbourhood of the zeros of ω . The neighbourhood of a zero of order k is isometric to a neighbourhood of the branching point of a $(k+1)$ -covering of a Euclidean disk.

The moduli space of all translation surfaces of a given genus g decomposes into strata. We focus only on those translation surfaces whose area is 1. Given a partition of $2g - 2 = \sum_{i=1}^{\ell} m_i$, the stratum $\mathcal{H}_g(m_1, \dots, m_\ell)$ is the space of all unit-area translation surfaces of genus g with ℓ zeros of order m_1, \dots, m_ℓ .

A *saddle connection* in (X, ω) is a geodesic segment that starts and ends in a zero and does not contain a zero in its interior. To any given (oriented) saddle connection γ on (X, ω) , we associate the vector $\int_\gamma \omega$ in \mathbb{C} , called the *holonomy vector* of the saddle connection. On a given translation surface, the number of saddle connections grows quadratically in their length [Mas88, Mas90].

A saddle connection can also be thought of as an element of the relative homology group of (X, ω) relative to the set of zeros. If \mathcal{B} is a set of saddle connections that form a basis for the relative homology, then the holonomy vectors of the elements of \mathcal{B} determine (X, ω) . For every (X, ω) and \mathcal{B} , there is a neighbourhood U of (X, ω) in the moduli space such that for every (X', ω') in U , all elements of \mathcal{B} (thought of as elements in the relative homology group) can still be represented in (X', ω') as saddle connections. Then the set of holonomy vectors of saddle connections in \mathcal{B} give coordinates for translation surfaces in U (which in general do not have area 1). We refer to this set of holonomy vectors as *period coordinates* around (X, ω) .

The period coordinates give an embedding from U to $\mathbb{C}^{2g+\ell-1}$. The associated pullback measure, denoted μ , of Lebesgue measure on $\mathbb{C}^{2g+\ell-1}$, on the moduli space was studied by Masur [Mas82] and Veech [Vee82]. It also defines a measure on the stratum $\mathcal{H}_g(m_1, \dots, m_\ell)$ in the following way. For an open set $V \subseteq \mathcal{H}_g(m_1, \dots, m_\ell)$, its volume is defined by

$$\text{Vol}(V) := 2(2g + \ell - 1) \cdot \mu(\text{Cone}(V))$$

where $\text{Cone}(V)$ is the cone of V in the moduli space and $2(2g + \ell - 1)$ is the real dimension of the cone. We refer to this measure on $\mathcal{H}_g(m_1, \dots, m_\ell)$ when we write $\text{Vol}(\cdot)$ in the following discussion.

Calculating volumes of strata explicitly is difficult in general. However, using an algorithm of Eskin–Okounkov [EO01], Eskin–Zorich predicted in [EZ15] and Aggarwal confirmed in [Agg20, Theorem 1.4] that we have

$$(1) \quad \text{Vol}(\mathcal{H}_g(m_1, \dots, m_\ell)) = \frac{4}{\prod_{i=1}^{\ell} (m_i + 1)} \cdot \left(1 + O\left(\frac{1}{g}\right)\right).$$

Recall that strata are not necessarily connected, they can have up to three connected components [KZ03]. However, the strata that we consider in this article are all connected.

4. THE METHOD OF SIEGEL–VEECH CONSTANTS

A useful method to calculate the probability of having a saddle connection of a certain length is by Siegel–Veech constants. To do this, we fix a stratum \mathcal{H} of unit-area translation surfaces. For $X \in \mathcal{H}$, $V_{sc}(X)$ denotes the set of (holonomy vectors of) saddle connections on X . A configuration \mathcal{C} of a saddle connection includes the multiplicity of the saddle connection, the order of the zeros that the saddle connection connects, and whether the zeros are different or not. Let $V_{\mathcal{C}}(X) \subseteq V_{sc}(X)$ denote the subset of holonomy vectors whose saddle connections are in the configuration \mathcal{C} . For an integrable function $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ with compact support, we define the *Siegel–Veech transform* $\hat{f}_{\mathcal{C}} : \mathcal{H} \rightarrow \mathbb{R}$ via

$$\hat{f}_{\mathcal{C}}(X) = \sum_{v \in V_{\mathcal{C}}(X)} f(v).$$

Theorem 4.1 (Siegel–Veech formula [Vee98]). *Let \mathcal{H} be a connected stratum of unit-area translation surfaces and \mathcal{C} a configuration of saddle connections. Then there exists a constant $c(\mathcal{C}, \mathcal{H})$ such that for every integrable $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ with compact support,*

$$(2) \quad \mathbb{E}(\hat{f}_{\mathcal{C}}) = \frac{\int_{\mathcal{H}} \hat{f}_{\mathcal{C}}}{\text{Vol}(\mathcal{H})} = c(\mathcal{C}, \mathcal{H}) \cdot \int_{\mathbb{R}^2} f.$$

The integral on the left is integration with respect to the normalized Lebesgue measure on \mathcal{H} , and the integral on the right is with respect to Lebesgue measure on \mathbb{R}^2 .

The constant $c(\mathcal{C}, \mathcal{H})$ is called the *Siegel–Veech constant* of the configuration and the stratum. It does not only appear when studying the average number of saddle connections for translation surfaces in a given stratum but also when counting saddle connections on a fixed translation surface of this stratum.

Theorem 4.2 (Quadratic growth of number of saddle connections [EM01]). *Let \mathcal{H} , \mathcal{C} , and $c(\mathcal{C}, \mathcal{H})$ as in Theorem 4.1. Then for almost every $(X, \omega) \in \mathcal{H}$, we have*

$$\lim_{T \rightarrow \infty} \frac{|V_{\mathcal{C}}(X) \cap B(0, T)|}{\pi T^2} = c(\mathcal{C}, \mathcal{H}).$$

Theorem 4.1 allows us to estimate expected values of various quantities on a stratum in terms of Siegel–Veech constants. Eskin–Masur–Zorich developed in [EMZ03] a method for computing Siegel–Veech constants using combinatorial constructions and surgeries as well as the Siegel–Veech formula above. Based on their recursive formulas, asymptotics of Siegel–Veech constants for large genus have been more recently calculated by many different authors, e.g. [CMZ18, Sau18, Agg20, Agg19, CMSZ20]. We recall now the Siegel–Veech constants for the configurations that are relevant for our proofs.

Let $\mathcal{C}_{m_1 \leftrightarrow m_2}$ be the configuration of a single saddle connection connecting a fixed zero of order m_1 to a fixed, different zero of order m_2 . Then the corresponding Siegel–Veech constant is [Agg20, Corollary 1], [Agg19, Theorem 1.2]

$$(3) \quad c_{m_1 \leftrightarrow m_2} = (m_1 + 1)(m_2 + 1) \cdot \left(1 + O\left(\frac{1}{g}\right)\right).$$

Let $\mathcal{C}_{1 \heartsuit 1}$ be the configuration of two homologous saddle connections connecting a fixed zero of order 1 to a fixed, different zero of order 1. Then the corresponding Siegel–Veech constant is [Agg19, Proposition 3.4]

$$(4) \quad c_{1 \heartsuit 1} = O\left(\frac{1}{g^2}\right).$$

Let $\mathcal{C}_{\circlearrowleft m}^{\text{non-hom}}$ be the configuration in $\mathcal{H}_g(m, 1, \dots, 1)$ of a saddle connection of multiplicity 1, connecting a fixed zero of order m to itself. Note that such a saddle connection is a loop which could either bound a cylinder or not, giving different cases for the formula for the Siegel–Veech constant. A combination of [Agg20, Corollary 3, 4, and 5] yields that the Siegel–Veech constant is

$$\begin{aligned} c_{\circlearrowleft m}^{\text{non-hom}} &= \frac{(m+1)(m-1)}{2} \left(1 + O\left(\frac{1}{g}\right)\right) + (m+1) \cdot \left(1 + O\left(\frac{1}{g}\right)\right) + O\left(\frac{1}{g}\right) \\ &= \frac{(m+1)^2}{2} \left(1 + O\left(\frac{1}{g}\right)\right). \end{aligned}$$

We will also need an estimate for $c_{\circlearrowleft m}$ where we allow saddle connections of higher multiplicity in $\mathcal{H}_g(m, 1, \dots, 1)$ as well as in $\mathcal{H}_g(m, 2, \dots, 2, 1, \dots, 1)$. With the formula

from [EMZ03, Section 13], this can be estimated as in the proof of [Agg20, Corollary 3, 4, and 5] to be

$$(5) \quad c_{\mathbb{O}m} = (m + 1)^2 \cdot O(1).$$

A detailed calculation of this Siegel–Veech constant can be found in [Val, Appendix 1].

5. COLLAPSING AND OPENING UP ZEROS

Our calculations depend heavily on the estimation of volumes of subsets of strata which contain only translation surfaces with short saddle connections. For this, we vary a construction from [EMZ03, Sections 8.1 and 8.2] that collapses a saddle connection between two different zeros of order 1. We describe here the construction and its inverse.

Let (X, ω) be a translation surface with two zeros v_1 and v_2 of order 1 that are connected by a saddle connection α of length δ . We assume for the description of the construction that α is horizontal. There are four horizontal separatrices starting in v_2 , one of them being α . Let γ be a geodesic segment of length δ on the horizontal separatrix starting in v_2 which forms an angle of 2π with α . We call γ the *ghost double* of α . Note that γ might not always exist and we will discuss in detail later what to do when it does not exist.

We now cut open along α and γ , obtaining two copies of α and γ each in the metric completion of $X \setminus (\alpha \cup \gamma)$. The upper copy of α we call α_1 and the lower copy α_2 , correspondingly for γ . Then we reglue α_1 with γ_1 and reglue α_2 with γ_2 as in Figure 1.

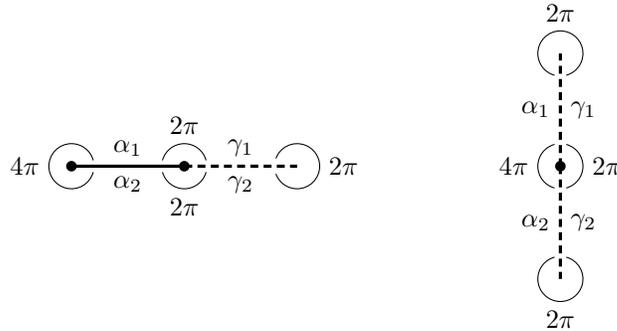


FIGURE 1. Collapsing a zero into another zero.

Note that the two points corresponding to v_2 under this cutting-and-gluing are regular points and that v_1 turns into a new zero of order 2. Therefore we call this construction *collapsing the zero v_2 into the zero v_1 (along α)* and the inverse of it *opening up a zero of order 2*. The data for opening up a zero of order 2 is the holonomy vector of the saddle connection α and a combinatorial information arising from choosing two out of the three geodesic segments with the given holonomy vector starting in the given zero of order 2.

We consider now $\mathcal{H}_g = \mathcal{H}_g(1, 1, \dots, 1)$, the principle stratum of unit-area translation surfaces of genus g with $2g - 2$ zeros of order 1. The process of collapsing a zero into another zero defines a map

$$(6) \quad \mathcal{H}^\epsilon \times \{\pm 1\} \rightarrow \mathcal{H}_g(2, 1, \dots, 1) \times M \times \mathbb{D}_\epsilon$$

where \mathcal{H}^ϵ is the subset of \mathcal{H}_g of translation surfaces with exactly one saddle connection of length at most ϵ and the choice of 1 or -1 indicates the orientation of this saddle connection. Furthermore, M is the set of possible combinatorial data of size $(2g - 2)(2g - 3)$ which records the names of the zeros that were connected by the saddle connection, and \mathbb{D}_ϵ is

the disk of radius ϵ in \mathbb{R}^2 . This map is well-defined and its inverse, defined by opening up the unique zero of order 2, is also well-defined up to the three possible choices of two geodesic segments. Hence, the map is 3-to-1. Furthermore, the map is locally measure-preserving: Let $(X, \omega) \in \mathcal{H}^\epsilon$ and \mathcal{B} be a basis of relative homology for (X, ω) which contains the unique shortest saddle connection α . Then \mathcal{B} defines period coordinates around (X, ω) and the image of $\mathcal{B} \setminus \alpha$ defines period coordinates around the translation surface in the image of (X, ω) whereas the holonomy vector of α defines coordinates for \mathbb{D}_ϵ . As the real dimensions of domain and image are both $2(4g-3)$ and the lengths of the vectors in a basis of domain and image coincide, also the measures of a neighbourhood of (X, ω) and its image coincide; see [EMZ03, Section 8.2] for more details.

If we consider several saddle connections which do not intersect (including not sharing a zero as end point) and whose ghost doubles do not intersect each other nor the saddle connections, we can collapse all pairs of zeros simultaneously, see Section 7. For intersecting saddle connections, this is in general not true. However, for a specific configuration, we can use the following simultaneous collapsing procedure.

Let (X, ω) be a translation surface with v_1, v_2 , and v_3 three zeros of order 1, α a saddle connection between v_1 and v_2 of length δ_1 , and β a saddle connection between v_2 and v_3 of length δ_2 . To collapse v_1 and v_3 into v_2 , we consider a ghost double of α , starting in v_1 and forming angle 2π with α , and a ghost double of β , starting in v_3 and forming angle 2π with β . Note that we have to make sure that neither of the ghost doubles and saddle connections intersect. The cases, in which they do intersect, we again exclude later by other volume estimations.

We cut open along α, β , and their ghost doubles and reglue as indicated in Figure 2. Then v_1 and v_3 turn into regular points and v_2 turns into a zero of order 3.

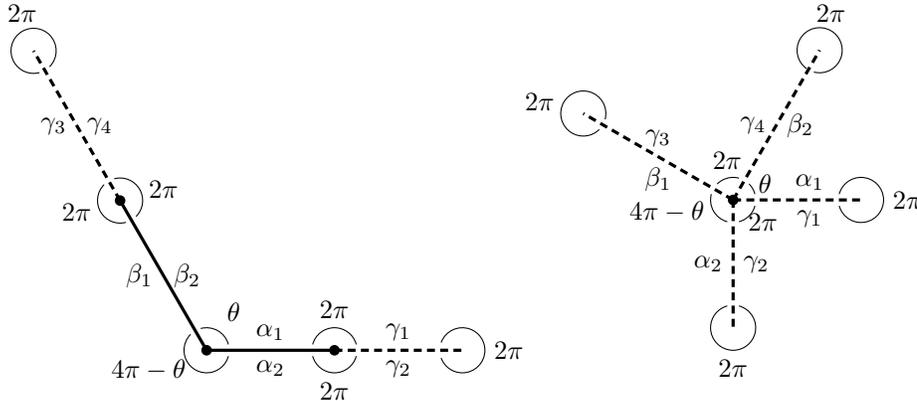


FIGURE 2. Collapsing two zeros into a third one, along two saddle connections.

6. THE GENERIC AND THE NON-GENERIC SUBSETS OF \mathcal{H}_g

Our strategy to prove Theorem 1.1 is to show the statement first for a generic subset $\mathcal{H}_{\text{Gen}} \subseteq \mathcal{H}_g$. Then we show that the proportion of the measure of the non-generic set \mathcal{H}_{NG} in \mathcal{H}_g goes to zero as $g \rightarrow \infty$.

Definition 6.1. Fix $0 < a_1 < b_1 \leq a_2 < b_2 \leq \dots \leq a_k < b_k \in \mathbb{R}_+$ and let $B := \max_i b_i$.

We define the following subsets of \mathcal{H}_g

$$\mathcal{H}_\circ := \{(X, \omega) \in \mathcal{H}_g : X \text{ contains a loop of length at most } 6B/g\}$$

$$\mathcal{H}_\lambda := \left\{ \begin{array}{l} (X, \omega) \in \mathcal{H}_g : X \text{ contains a saddle connection of length at most } B/g \text{ and} \\ \text{a saddle connection of length at most } 2B/g \text{ that share a zero} \end{array} \right\}$$

where a *loop* is a saddle connection that connects a zero to itself.

Furthermore, we define the *non-generic subset* of \mathcal{H}_g to be $\mathcal{H}_{\text{NG}} = \mathcal{H}_\circ \cup \mathcal{H}_\lambda$ and the *generic subset* to be the complement of the non-generic set: $\mathcal{H}_{\text{Gen}} = \mathcal{H}_g \setminus \mathcal{H}_{\text{NG}}$.

The next proposition justifies calling the subsets generic and non-generic.

Proposition 6.2. *Let \mathcal{H}_{Gen} as in Definition 6.1. Then*

$$\frac{\text{Vol}(\mathcal{H}_{\text{Gen}})}{\text{Vol}(\mathcal{H}_g)} \rightarrow 1 \quad \text{as } g \rightarrow \infty.$$

More specifically,

$$\frac{\text{Vol}(\mathcal{H}_{\text{NG}})}{\text{Vol}(\mathcal{H}_g)} = O\left(\frac{1}{g}\right)$$

where the implied constant does not depend on g .

Proof. We first use the method of Siegel–Veech constants to obtain an estimate on the volume of \mathcal{H}_\circ . For this, let $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ be the indicator function of the ball of radius $6B/g$ centered at 0. Choose $\mathcal{C}_{\circ 1}$ to be the configuration of a saddle connections which forms a loop, starting at a zero of order 1.

Applying Theorem 4.1 and the Siegel–Veech constant from Equation (5), we obtain

$$(7) \quad \text{Vol}(\mathcal{H}_\circ) = \int_{\mathcal{H}_g} \mathbb{1} \leq \int_{\mathcal{H}_g} \hat{f}_{\mathcal{C}_{\circ 1}} = (2g-2) \cdot c_{\circ 1} \cdot \int_{\mathbb{R}^2} f \cdot \text{Vol}(\mathcal{H}_g) = O\left(\frac{B^2}{g}\right) \cdot \text{Vol}(\mathcal{H}_g),$$

where the factor $2g-2$ corresponds to the number of possible choices for the zero at which the loop is based.

We will later in the proof need similar estimates as (7) for the following subset of \mathcal{H}_g

$$\mathcal{H}_\infty := \left\{ \begin{array}{l} (X, \omega) \in \mathcal{H}_g : X \text{ contains a pair of homologous saddle connections} \\ \text{of length at most } B/g \text{ between different zeros} \end{array} \right\}$$

and the following two subsets of $\mathcal{H}_g(2, 1, \dots, 1)$

$$\mathcal{H}_{\circ 2} := \left\{ \begin{array}{l} (X, \omega) \in \mathcal{H}_g(2, 1, \dots, 1) : X \text{ contains a loop of length at most } 6B/g \\ \text{starting at the zero of order 2} \end{array} \right\}$$

$$\mathcal{H}_{2 \leftrightarrow 1} := \left\{ \begin{array}{l} (X, \omega) \in \mathcal{H}_g(2, 1, \dots, 1) : X \text{ contains a saddle connection of length at most } 3B/g \\ \text{from the zero of order 2 to a zero of order 1} \end{array} \right\}.$$

Analogously to (7), we obtain with the Siegel–Veech constant from (4)

$$(8) \quad \text{Vol}(\mathcal{H}_\infty) = O\left(\frac{B^2}{g^2}\right) \cdot \text{Vol}(\mathcal{H}_g),$$

with the Siegel–Veech constant from (5) for $m=2$

$$(9) \quad \text{Vol}(\mathcal{H}_{\circ 2}) = O\left(\frac{B^2}{g^2}\right) \cdot \text{Vol}(\mathcal{H}_g(2, 1, \dots, 1)),$$

and with the Siegel–Veech constant from (3) for $m_1 = 2$ and $m_2 = 1$

$$(10) \quad \text{Vol}(\mathcal{H}_{2 \leftarrow 1}) = O\left(\frac{B^2}{g}\right) \cdot \text{Vol}(\mathcal{H}_g(2, 1, \dots, 1)).$$

To estimate the volume of \mathcal{H}_Λ , the idea is to simultaneously collapse the two saddle connections that share a zero. We introduce another subset of \mathcal{H}_g which contains some of the cases for which the collapsing is not possible. Let

$$\mathcal{H}_\triangleright := \{(X, \omega) \in \mathcal{H}_g : X \text{ contains a closed geodesic of length at most } 6B/g\}.$$

Note that a closed geodesic on a translation surface in $\mathcal{H}_\triangleright$ is either a chain of saddle connections with angles at least π between them or is contained in a cylinder. In the latter case, we can represent the closed geodesic as a boundary curve of the cylinder and hence we have a chain of saddle connections again.

We already have an upper bound for the volume of $\mathcal{H}_\circlearrowleft$. To determine an upper bound for the volume of $\mathcal{H}_{\text{NG}} = \mathcal{H}_\circlearrowleft \cup \mathcal{H}_\Lambda$, we consider now $\mathcal{H}_\Lambda \setminus (\mathcal{H}_\circlearrowleft \cup \mathcal{H}_\triangleright)$, $\mathcal{H}_\triangleright \setminus (\mathcal{H}_\circlearrowleft \cup \mathcal{H}_\infty)$ and \mathcal{H}_∞ separately.

First, for a translation surface in $\mathcal{H}_\Lambda \setminus (\mathcal{H}_\circlearrowleft \cup \mathcal{H}_\triangleright)$, let α be the shortest saddle connection that shares a zero with a saddle connection of length at most $2B/g$. Furthermore, let β be the shortest saddle connection that α shares a zero with. Then we can collapse two zeros along α and β into the shared zero. As the length of α is at most B/g and the length of β is at most $2B/g$, this gives us the following locally measure-preserving map:

$$\mathcal{H}_\Lambda \setminus (\mathcal{H}_\circlearrowleft \cup \mathcal{H}_\triangleright) \rightarrow M_\Lambda \times \mathcal{H}_g(3, 1, \dots, 1) \times \mathbb{D}_{B/g} \times \mathbb{D}_{2B/g}$$

Note that this map is well-defined on a subset of full measure: The saddle connections α and β are uniquely determined except for a subset of measure 0. And if α and β or their ghost doubles would intersect, α would be a loop or we would have a closed geodesic of length at most $6B/g$ that shares a zero with α , so the translation surface would be in $\mathcal{H}_\circlearrowleft \cup \mathcal{H}_\triangleright$. The set M_Λ here expresses the possible choices of the three zeros involved and hence is of order g^3 . We can define an inverse of the map up to a finite choice of the geodesic segments which are to be cut open. Hence, the map is finite-to-1 and we obtain with Equation (1)

$$\begin{aligned} \frac{\text{Vol}(\mathcal{H}_\Lambda \setminus (\mathcal{H}_\circlearrowleft \cup \mathcal{H}_\triangleright))}{\text{Vol}(\mathcal{H}_g)} &= O(1) \cdot |M_\Lambda| \cdot \frac{\text{Vol}(\mathcal{H}_g(3, 1, \dots, 1))}{\text{Vol}(\mathcal{H}_g)} \cdot \text{Vol}(\mathbb{D}_{B/g} \times \mathbb{D}_{2B/g}) \\ &= O\left(g^3 \cdot \frac{4}{2^{2g-5} \cdot 4} \cdot \frac{2^{2g-2}}{4} \cdot \frac{B^4}{g^4}\right) = B^4 \cdot O\left(\frac{1}{g}\right). \end{aligned}$$

For the set $\mathcal{H}_\triangleright \setminus (\mathcal{H}_\circlearrowleft \cup \mathcal{H}_\infty)$, we approach the volume estimate in two steps: For a translation surface in $\mathcal{H}_\triangleright \setminus (\mathcal{H}_\circlearrowleft \cup \mathcal{H}_\infty)$, choose the shortest closed geodesic. Let α be the shortest saddle connection on this closed geodesic. Then α is not a loop and we can collapse a zero along α into another zero such that we obtain a closed geodesic of length at most $6B/g$ containing a zero of order 2. This gives us the following finite-to-1, locally measure-preserving map:

$$\mathcal{H}_\triangleright \setminus (\mathcal{H}_\circlearrowleft \cup \mathcal{H}_\infty) \rightarrow M_\triangleright \times (\mathcal{H}_{\circlearrowleft 2} \cup \mathcal{H}_{2 \leftarrow 1}) \times \mathbb{D}_{6B/g}$$

Note that this map is well-defined on a subset of full measure: The ghost double of α would not be defined only if there is a saddle connection of shorter length with the same direction (which only happens for a set of measure zero) or the ghost double defines a saddle connection which is homologous to α (which would mean that we are in \mathcal{H}_∞). Here, M_\triangleright is the combinatorial set of order g^2 that expresses the choices of zeros that are collapsed into

each other. We can estimate the volume of $\mathcal{H}_{\mathcal{O}_2} \cup \mathcal{H}_{2 \leftrightarrow 1}$ using Equations (9) and (10) as

$$\frac{\text{Vol}(\mathcal{H}_{\mathcal{O}_2} \cup \mathcal{H}_{2 \leftrightarrow 1})}{\text{Vol}(\mathcal{H}_g(2, 1, \dots, 1))} = O\left(\frac{B^2}{g^2}\right) + O\left(g \cdot \frac{B^2}{g^2}\right) = B^2 \cdot O\left(\frac{1}{g}\right).$$

Therefore, we obtain with Equation (1)

$$\frac{\text{Vol}(\mathcal{H}_{\mathcal{O}} \setminus (\mathcal{H}_{\mathcal{O}} \cup \mathcal{H}_{\infty}))}{\text{Vol}(\mathcal{H}_g)} = O(1) \cdot |M_{\mathcal{O}}| \cdot \frac{\text{Vol}(\mathcal{H}_{\mathcal{O}_2} \cup \mathcal{H}_{2 \leftrightarrow 1})}{\text{Vol}(\mathcal{H}_g)} \cdot \text{Vol}(\mathbb{D}_{6B/g}) = B^4 \cdot O\left(\frac{1}{g}\right).$$

Finally, with Equations (7) and (8), we have

$$\begin{aligned} \frac{\text{Vol}(\mathcal{H}_{\text{NG}})}{\text{Vol}(\mathcal{H}_g)} &\leq \frac{\text{Vol}(\mathcal{H}_{\mathcal{O}})}{\text{Vol}(\mathcal{H}_g)} + \frac{\text{Vol}(\mathcal{H}_{\Lambda} \setminus (\mathcal{H}_{\mathcal{O}} \cup \mathcal{H}_{\mathcal{O}}))}{\text{Vol}(\mathcal{H}_g)} + \frac{\text{Vol}(\mathcal{H}_{\mathcal{O}} \setminus (\mathcal{H}_{\mathcal{O}} \cup \mathcal{H}_{\infty}))}{\text{Vol}(\mathcal{H}_g)} + \frac{\text{Vol}(\mathcal{H}_{\infty})}{\text{Vol}(\mathcal{H}_g)} \\ &= B^2 \cdot O\left(\frac{1}{g}\right) + B^4 \cdot O\left(\frac{1}{g}\right) + B^4 \cdot O\left(\frac{1}{g}\right) + B^2 \cdot O\left(\frac{1}{g^2}\right) \\ &= \max\{B^2, B^4\} \cdot O\left(\frac{1}{g}\right). \end{aligned} \quad \square$$

We will also need a variation of Proposition 6.2 for strata different from the principal stratum, specifically for the strata to which translation surfaces belong after collapsing along K saddle connections.

Definition 6.3. Fix $K \in \mathbb{N}, K < g$ and let $\mathcal{H}' = \mathcal{H}_g(2, \dots, 2, 1, \dots, 1)$ be the stratum with K zeros of order 2 and $2g - 2 - 2K$ zeros of order 1.

Furthermore, fix $0 < a_1 < b_1 \leq a_2 < b_2 \leq \dots \leq a_k < b_k \in \mathbb{R}_+$ and let $B := \max_i b_i$.

We define the following subsets of \mathcal{H}' :

$$\begin{aligned} \mathcal{H}'_{\mathcal{O}_1} &:= \left\{ (X, \omega) \in \mathcal{H}' : \begin{array}{l} X \text{ contains a loop of length at most } 6B/g \\ \text{starting at a zero of order 1} \end{array} \right\} \\ \mathcal{H}'_{\Lambda} &:= \left\{ (X, \omega) \in \mathcal{H}' : \begin{array}{l} X \text{ contains a saddle connection of length at most } B/g \text{ and} \\ \text{a saddle connection of length at most } 2B/g \text{ that share a zero} \\ \text{where all zeros are of order 1} \end{array} \right\} \\ \mathcal{H}'_{\mathcal{O}_2} &:= \left\{ (X, \omega) \in \mathcal{H}' : \begin{array}{l} X \text{ contains a loop of length at most } 2B/g \\ \text{starting at a zero of order 2} \end{array} \right\} \\ \mathcal{H}'_{2 \leftrightarrow * } &:= \left\{ (X, \omega) \in \mathcal{H}' : \begin{array}{l} X \text{ contains a saddle connection of length at most } 2B/g \\ \text{from a zero of order 2 to another zero} \end{array} \right\} \end{aligned}$$

Similarly to before, we define the *non-generic subset* of \mathcal{H}' to be

$$\mathcal{H}'_{\text{NG}} = \mathcal{H}'_{\mathcal{O}_1} \cup \mathcal{H}'_{\Lambda} \cup \mathcal{H}'_{\mathcal{O}_2} \cup \mathcal{H}'_{2 \leftrightarrow *}$$

and the *generic subset* to be the complement of the non-generic set: $\mathcal{H}'_{\text{Gen}} = \mathcal{H}' \setminus \mathcal{H}'_{\text{NG}}$.

We can prove as in Proposition 6.2 that the volume of \mathcal{H}'_{NG} is small:

Proposition 6.4. For fixed $K \in \mathbb{N}$, we have

$$\frac{\text{Vol}(\mathcal{H}'_{\text{Gen}})}{\text{Vol}(\mathcal{H}')} \rightarrow 1 \quad \text{as } g \rightarrow \infty.$$

More specifically,

$$\frac{\text{Vol}(\mathcal{H}'_{\text{NG}})}{\text{Vol}(\mathcal{H}')} = O\left(\frac{K}{g}\right)$$

where the implied constant does not depend on K or g .

Proof. Similar to Equations (9) and (10), we can use Equations (5) and (3) to obtain

$$\frac{\text{Vol}(\mathcal{H}'_{\mathcal{O}_2} \cup \mathcal{H}'_{2 \leftrightarrow *})}{\text{Vol}(\mathcal{H}')} = 9B^2 \cdot O\left(K \cdot \frac{1}{g^2}\right) + 6B^2 \cdot O\left(Kg \cdot \frac{1}{g^2}\right) + 9B^2 \cdot O\left(K^2 \cdot \frac{1}{g^2}\right),$$

where the third (and first) term is dominated by the second as $K < g$.

Furthermore, as in the proof of Proposition 6.2, we can show that

$$\frac{\text{Vol}(\mathcal{H}'_{\lambda} \cup \mathcal{H}'_{\mathcal{O}_1})}{\text{Vol}(\mathcal{H}')} = B^4 \cdot K \cdot O\left(\frac{1}{g}\right) + B^2 \cdot O\left(\frac{1}{g^2}\right).$$

Note that the additional factor of K comes from the fact that we now have $K + 1$ zeros of order 2 in $\mathcal{H}'_{\mathcal{O}_2} \cup \mathcal{H}'_{2 \leftrightarrow *}$ instead of only one in $\mathcal{H}_{\mathcal{O}_2} \cup \mathcal{H}_{2 \leftrightarrow 1}$.

In summary, we can deduce that

$$\frac{\text{Vol}(\mathcal{H}'_{\text{NG}})}{\text{Vol}(\mathcal{H}')} = \max\{B^2, B^4\} \cdot K \cdot O\left(\frac{1}{g}\right) \quad \text{and} \quad \frac{\text{Vol}(\mathcal{H}'_{\text{Gen}})}{\text{Vol}(\mathcal{H}')} \rightarrow 1. \quad \square$$

7. SIMULTANEOUS COLLAPSING

As before, fix $0 < a_1 < b_1 \leq a_2 < b_2 \leq \dots \leq a_k < b_k \in \mathbb{R}_+$ and let $B := \max_i b_i$. Furthermore, fix $r_1, \dots, r_k \in \mathbb{N}$ and define $K := \sum_{i=1}^k r_i$. To apply Theorem 2.1, we have to determine

$$\int_{\mathcal{H}_{\text{Gen}}} (N_{g, [a_1, b_1]})_{r_1} \cdots (N_{g, [a_k, b_k]})_{r_k}.$$

Instead of calculating this integral directly, we replace \mathcal{H}_{Gen} by a feasible cover $\widehat{\mathcal{H}}$, such that we can replace counting saddle connections on elements of \mathcal{H}_{Gen} by calculating the volume of $\widehat{\mathcal{H}}$.

For this, let $\widehat{\mathcal{H}}$ be the set of tuples (X, ω, Γ) where $(X, \omega) \in \mathcal{H}_{\text{Gen}}$, $\Gamma = (\Gamma_1, \dots, \Gamma_k)$ and $\Gamma_i = (\alpha_{i,1}, \dots, \alpha_{i,r_i})$ is an ordered list of disjoint oriented saddle connections with lengths in the interval $[a_i/g, b_i/g]$ for every $i \in \{1, \dots, k\}$. We sometimes think of Γ as a set and refer to a saddle connection $\alpha \in \Gamma$, which means α belongs to some $\Gamma_i \in \Gamma$.

Note that none of the saddle connections in Γ is a loop and that no two saddle connections in Γ or their ghost doubles intersect each other: If the latter happened, there would exist another saddle connection of length at most $2B/g$ that shares a zero with one of the saddle connections in Γ , so (X, ω) would be in \mathcal{H}_{NG} . Therefore, we can simultaneously collapse all pairs of zeros that are connected by saddle connections in Γ to obtain a map

$$\Phi: \widehat{\mathcal{H}} \rightarrow M_{r_1, \dots, r_k} \times \mathcal{H}' \times \prod_{i=1}^k A_{a_i, b_i}^{r_i}.$$

Here $\mathcal{H}' = \mathcal{H}_g(2, \dots, 2, 1, \dots, 1)$ is the stratum of translation surfaces of genus g with K zeros of order 2 and the rest of order 1. The set M_{r_1, \dots, r_k} consists of the possible combinatorial data needed to record the labels of the (ordered) pairs of zeros of order 1 in $\widehat{\mathcal{H}}$ which give rise to zeros of order 2 in \mathcal{H}' . Finally, the set $A_{a_i, b_i} \subseteq \mathbb{R}^2$ is an annulus in \mathbb{R}^2 centered at the origin with inner radius a_i/g and outer radius b_i/g and the notation $A_{a_i, b_i}^{r_i}$ means that we take r_i copies of this annulus.

With the same arguments as for the map in (6), the map Φ is 3^K -to-1 and locally measure-preserving. We use this now to estimate the volume of $\widehat{\mathcal{H}}$.

Lemma 7.1. *For $\widehat{\mathcal{H}}$ as above, we have*

$$\text{Vol}(\widehat{\mathcal{H}}) = 3^K \cdot \text{Vol}(\mathcal{H}'_{\text{Gen}}) \cdot \left(1 + K \cdot O\left(\frac{1}{g}\right)\right) \cdot \prod_{i=1}^k \left(\frac{\pi(b_i^2 - a_i^2)}{g^2}\right)^{r_i} \cdot |M_{r_1, \dots, r_k}|$$

where the implied constant does not depend on K or g .

Proof. For any $(X', \omega') \in \mathcal{H}'_{\text{Gen}}$, any element of M_{r_1, \dots, r_k} and any K -tuple of vectors in $\prod_{i=1}^k A_{a_i, b_i}^{r_i}$, the K zeros of order 2 can be opened up (in 3^K different ways) in the direction of vectors in the chosen tuple and labelled according to the element of M_{r_1, \dots, r_k} to obtain (X, ω) in \mathcal{H}_{Gen} . Furthermore, by choosing the oriented saddle connections in X obtained from opening up zeros of order 2 as the elements of the tuples Γ_i , we obtain a point in $\widehat{\mathcal{H}}$. That is, the chosen point (X', ω') together with the further data is in the image of Φ .

Therefore, we have

$$\mathcal{H}'_{\text{Gen}} \times \prod A_{a_i, b_i}^{r_i} \times M_{r_1, \dots, r_k} \subseteq \Phi(\widehat{\mathcal{H}}) \subseteq \mathcal{H}' \times \prod A_{a_i, b_i}^{r_i} \times M_{r_1, \dots, r_k}.$$

As the map Φ is 3^K -to-1, locally measure-preserving and nearly onto as just described and with the estimates on the volume of the complement of $\mathcal{H}'_{\text{Gen}}$ from Proposition 6.4, we obtain the claimed volume estimate. \square

8. THE LENGTH SPECTRUM

In this section, we check the assumptions of Theorem 2.1 for the set \mathcal{H}_{Gen} . As before, fix $0 < a_1 < b_1 \leq a_2 < b_2 \leq \dots \leq a_k < b_k \in \mathbb{R}_+$. For positive integers $r_1, \dots, r_k \in \mathbb{N}$, define

$$Y_{g, r_1, \dots, r_k} := (N_{g, [a_1, b_1]})_{r_1} \dots (N_{g, [a_k, b_k]})_{r_k} : \mathcal{H}_g \rightarrow \mathbb{N}_0.$$

Then Y_{g, r_1, \dots, r_k} counts the number of (ordered) lists of length k where the i -th item is an ordered r_i -tuple of saddle connections with lengths in $[a_i/g, b_i/g]$ on a surface $(X, \omega) \in \mathcal{H}_g$.

Define

$$\mathbb{E}_{\text{Gen}}(\cdot) = \frac{1}{\text{Vol}(\mathcal{H}_{\text{Gen}})} \int_{\mathcal{H}_{\text{Gen}}} \cdot.$$

Proposition 8.1. *For fixed $0 < a_1 < b_1 \leq \dots \leq a_k < b_k \in \mathbb{R}_+$, we have*

$$\lim_{g \rightarrow \infty} \mathbb{E}_{\text{Gen}}(Y_{g, r_1, \dots, r_k}) = \prod_{i=1}^k \lambda_{[a_i, b_i]}^{r_i} \quad \text{with} \quad \lambda_{[a_i, b_i]} = 8\pi(b_i^2 - a_i^2).$$

Proof. Let $K := \sum_{i=1}^k r_i$ as before. For every $(X, \omega) \in \mathcal{H}_{\text{Gen}}$, every suitable set of K saddle connections gives rise to 2^K elements in $\widehat{\mathcal{H}}$ with different choices of orientations for the saddle connections in Γ . Hence we have

$$\mathbb{E}_{\text{Gen}}(Y_{g, r_1, \dots, r_k}) = \frac{1}{\text{Vol} \mathcal{H}_{\text{Gen}}} \cdot \frac{1}{2^K} \int_{\widehat{\mathcal{H}}} \mathbb{1}.$$

The volume estimate from Lemma 7.1 implies

$$\int_{\widehat{\mathcal{H}}} \mathbb{1} = 3^K \cdot \text{Vol}(\mathcal{H}'_{\text{Gen}}) \cdot \prod_{i=1}^k \left(\frac{\pi(b_i^2 - a_i^2)}{g^2} \right)^{r_i} \cdot |M_{r_1, \dots, r_k}| \cdot \left(1 + K \cdot O\left(\frac{1}{g}\right) \right)$$

with the notation as in Section 7.

An element in M_{r_1, \dots, r_k} is a way to choose $2K$ zeros (with labels) successively out of a set of $2g - 2$ zeros with labels, hence

$$|M_{r_1, \dots, r_k}| = \frac{(2g - 2)!}{(2g - 2 - 2K)!}.$$

Combining these formulas, we get

$$\mathbb{E}_{\text{Gen}}(Y_{g, r_1, \dots, r_k}) = \frac{\text{Vol}(\mathcal{H}'_{\text{Gen}})}{\text{Vol}(\mathcal{H}_{\text{Gen}})} \cdot \frac{3^K}{2^K} \cdot \frac{(2g - 2)!}{(2g - 2 - 2K)!} \cdot \prod_{i=1}^k \left(\frac{\pi(b_i^2 - a_i^2)}{g^2} \right)^{r_i} \cdot \left(1 + K \cdot O\left(\frac{1}{g}\right) \right).$$

Furthermore, we have from Propositions 6.2 and 6.4 and Equation (1) that

$$\lim_{g \rightarrow \infty} \frac{\text{Vol}(\mathcal{H}'_{\text{Gen}})}{\text{Vol}(\mathcal{H}_{\text{Gen}})} = \lim_{g \rightarrow \infty} \frac{\text{Vol}(\mathcal{H}'_g)}{\text{Vol}(\mathcal{H}_g)} = \frac{4}{3^K \cdot 2^{2g-2-2K}} \cdot \frac{2^{2g-2}}{4} = \frac{2^{2K}}{3^K} = \left(\frac{4}{3}\right)^K.$$

Taking the limit of $\mathbb{E}_{\text{Gen}}(Y_{g,r_1,\dots,r_k})$ for $g \rightarrow \infty$ and using $K = \sum_{i=1}^k r_i$ yields

$$\begin{aligned} \lim_{g \rightarrow \infty} \mathbb{E}_{\text{Gen}}(Y_{g,r_1,\dots,r_k}) &= \lim_{g \rightarrow \infty} \left(\frac{4}{3}\right)^K \cdot \frac{3^K}{2^K} \cdot (2g)^{2K} \cdot \frac{1}{g^{2K}} \cdot \prod_{i=1}^k (\pi(b_i^2 - a_i^2))^{r_i} \\ &= \prod_{i=1}^k (8\pi(b_i^2 - a_i^2))^{r_i}. \end{aligned} \quad \square$$

9. FROM THE GENERIC SET TO THE WHOLE STRATUM

Let $\mathbb{P}(\bullet)$ denote the probability of an event in \mathcal{H}_g and let $\mathbb{P}_{\text{Gen}}(\bullet)$ denote the probability of an event in \mathcal{H}_{Gen} . For a random variable $N: \mathcal{H}_g \rightarrow \mathbb{N}_0$ and $k \in \mathbb{N}$, we have

$$\mathbb{P}_{\text{Gen}}(N = k) = \frac{\text{Vol}\left\{(X, \omega) \in \mathcal{H}_{\text{Gen}} \mid N(X, \omega) = k\right\}}{\text{Vol}(\mathcal{H}_{\text{Gen}})}.$$

To compare $\mathbb{P}(\bullet)$ and $\mathbb{P}_{\text{Gen}}(\bullet)$, note that

$$\begin{aligned} \text{Vol}\left\{(X, \omega) \in \mathcal{H}_g \mid N(X, \omega) = k\right\} - \text{Vol}(\mathcal{H}_{\text{NG}}) &\leq \text{Vol}\left\{(X, \omega) \in \mathcal{H}_{\text{Gen}} \mid N(X, \omega) = k\right\} \\ &\leq \text{Vol}\left\{(X, \omega) \in \mathcal{H}_g \mid N(X, \omega) = k\right\} \end{aligned}$$

From Proposition 6.2, we know that

$$\frac{\text{Vol}(\mathcal{H}_{\text{Gen}})}{\text{Vol}(\mathcal{H}_g)} \rightarrow 1 \quad \text{and} \quad \frac{\text{Vol}(\mathcal{H}_{\text{NG}})}{\text{Vol}(\mathcal{H}_g)} \rightarrow 0 \quad \text{as} \quad g \rightarrow \infty.$$

Hence, if $\lim_{g \rightarrow \infty} \mathbb{P}_{\text{Gen}}(N = k)$ exists, so does $\lim_{g \rightarrow \infty} \mathbb{P}(N = k)$ and the limits are equal. The same holds for any event involving multiple random variables.

In the case discussed here, from Proposition 8.1 and Theorem 2.1, it follows that

$$\lim_{g \rightarrow \infty} \mathbb{P}_{\text{Gen}}(N_{g,[a_1,b_1]} = n_1, \dots, N_{g,[a_k,b_k]} = n_k) = \prod_{i=1}^k \frac{\lambda_{[a_i,b_i]}^{n_i} e^{-\lambda_{[a_i,b_i]}}}{n_i!}$$

Therefore,

$$\lim_{g \rightarrow \infty} \mathbb{P}(N_{g,[a_1,b_1]} = n_1, \dots, N_{g,[a_k,b_k]} = n_k) = \prod_{i=1}^k \frac{\lambda_{[a_i,b_i]}^{n_i} e^{-\lambda_{[a_i,b_i]}}}{n_i!}$$

This finishes the proof of Theorem 1.1.

10. PROOF OF THEOREMS 1.2 AND 1.3

In this section, we outline how to prove the two versions of the main theorem for other strata than the principal stratum.

We start with $\mathcal{H}_g(m, \dots, m)$ where m is a fixed odd number that divides $2g - 2$. To adapt the proof of Proposition 6.2 to this situation, we replace all the zeros of order 1 by zeros of order m . The collapsing procedure still works when considering all of the potential m ghost doubles simultaneously. As m is fixed, the values of the Siegel–Veech constants and the combinatorial constants are changed but not the order of g in these terms. Hence, the statement of Proposition 6.2 and the analogous version of Proposition 6.4 is also true for $\mathcal{H}_g(m, \dots, m)$.

The map Φ from Section 7 is replaced by the map

$$\widehat{\mathcal{H}} \rightarrow M_{r_1, \dots, r_k} \times \mathcal{H}'' \times \prod_{i=1}^k A_{a_i, b_i}^{r_i},$$

where $\mathcal{H}'' = \mathcal{H}_g(2m, \dots, 2m, m, \dots, m)$ is the stratum of translation surfaces of genus g with K zeros of order $2m$ and the rest of order m . Note that this map is $(2m+1)^{K-1}$ -to-1 which changes the volume of $\widehat{\mathcal{H}}$ to

$$\text{Vol}(\widehat{\mathcal{H}}) = (2m+1)^K \cdot \text{Vol}(\mathcal{H}_{\text{Gen}}'') \cdot \left(1 + K \cdot O\left(\frac{1}{g}\right)\right) \cdot \prod_{i=1}^k \left(\frac{\pi(b_i^2 - a_i^2)}{g^2}\right)^{r_i} \cdot |M_{r_1, \dots, r_k}|.$$

With this, we can now do the same calculation as in Proposition 8.1. In particular, we have

$$|M_{r_1, \dots, r_k}| = \frac{\left(\frac{2g-2}{m}\right)!}{\left(\frac{2g-2}{m} - 2K\right)!}$$

and

$$\lim_{g \rightarrow \infty} \frac{\text{Vol}(\mathcal{H}_{\text{Gen}}'')}{\text{Vol}(\mathcal{H}_{\text{Gen}})} = \frac{4}{(2m+1)^K \cdot (m+1)^{\frac{2g-2}{m} - 2K}} \cdot \frac{(m+1)^{\frac{2g-2}{m}}}{4} = \left(\frac{(m+1)^2}{2m+1}\right)^K.$$

Hence with $K = \sum_{i=1}^k r_i$, we obtain

$$\begin{aligned} \lim_{g \rightarrow \infty} \mathbb{E}_{\text{Gen}}(Y_{g, r_1, \dots, r_k}) &= \lim_{g \rightarrow \infty} \left(\frac{(m+1)^2}{2m+1}\right)^K \cdot \frac{(2m+1)^K}{2^K} \cdot \left(\frac{2g-2}{m}\right)^{2K} \cdot \frac{1}{g^{2K}} \cdot \prod_{i=1}^k (\pi(b_i^2 - a_i^2))^{r_i} \\ &= \prod_{i=1}^k \left(\left(\frac{m+1}{m}\right)^2 2\pi(b_i^2 - a_i^2)\right)^{r_i} \end{aligned}$$

which proves Theorem 1.3.

For $\mathcal{H}_g(m_1, m_2, \dots, m_{\ell(g)}, 1, \dots, 1)$ where $m_i \leq f(g)$, $i = 1, \dots, \ell(g)$, we consider the following: We use Equation (3) to show that the volume proportion of translation surfaces with a saddle connection of length at most B/g from a non-simple zero to any other zero is at most of the order of

$$(f(g)+1)^2 \cdot \ell(g) \cdot g \cdot \left(\frac{B}{g}\right)^2 = O\left(\frac{f(g)^2 \cdot \ell(g)}{g}\right).$$

Therefore, if $\frac{f(g)^2 \cdot \ell(g)}{g} \rightarrow 0$ for $g \rightarrow \infty$, the occurrence of these saddle connections is dominated by the occurrence of saddle connections from a simple zero to another simple zero that are used in the volume calculation in Proposition 8.1. Hence, the same reasoning as in the principal stratum works to obtain Theorem 1.2.

REFERENCES

- [Agg19] Amol Aggarwal. Large genus asymptotics for Siegel-Veech constants. *Geometric and Functional Analysis (GAFA)*, 29(5):1295–1324, 2019.
- [Agg20] Amol Aggarwal. Large Genus Asymptotics for Volumes of Strata of Abelian Differentials. *Journal of the American Mathematical Society*, 33(4):941–989, 2020. With an appendix by Anton Zorich.
- [Bol01] Béla Bollobás. *Random graphs.*, volume 73 of *Cambridge Studies in Advanced Mathematics*. 2nd ed. edition, 2001.
- [BRV24] Lewis Bowen, Kasra Rafi, and Hunter Vallejos. Benjamini–Schramm limits of high genus translation surfaces. In preparation, 2024.
- [CMS23] Dawei Chen, Martin Möller, and Adrien Sauvaget. Masur-Veech volumes and intersection theory: the principal strata of quadratic differentials. With an appendix by Gaëtan Borot, Alessandro Giacchetto and Danilo Lewanski. *Duke Mathematical Journal*, 172(9):1735–1779, 2023.

- [CMSZ20] Dawei Chen, Martin Möller, Adrien Sauvaget, and Don Zagier. Masur-Veech volumes and intersection theory on moduli spaces of abelian differentials. *Inventiones Mathematicae*, 222(1):283–373, 2020.
- [CMZ18] Dawei Chen, Martin Möller, and Don Zagier. Quasimodularity and large genus limits of Siegel-Veech constants. *Journal of the American Mathematical Society*, 31(4):1059–1163, 2018.
- [DGZZ23] Vincent Delecroix, Elise Goujard, Peter Zograf, and Anton Zorich. Higher genus meanders and Masur-Veech volumes. Preprint, arXiv:2304.02567 [math.GT] (2023), 2023.
- [EM01] Alex Eskin and Howard Masur. Asymptotic formulas on flat surfaces. *Ergodic Theory and Dynamical Systems*, 21(2):443–478, 2001.
- [EMZ03] Alex Eskin, Howard Masur, and Anton Zorich. Moduli spaces of Abelian differentials: the principal boundary, counting problems, and the Siegel-Veech constants. *Publications Mathématiques, Institut des Hautes Études Scientifiques*, 97:61–179, 2003.
- [EO01] Alex Eskin and Andrei Okounkov. Asymptotics of numbers of branched coverings of a torus and volumes of moduli spaces of holomorphic differentials. *Inventiones Mathematicae*, 145(1):59–103, 2001.
- [EZ15] Alex Eskin and Anton Zorich. Volumes of strata of abelian differentials and Siegel-Veech constants in large genera. *Arnold Mathematical Journal*, 1(4):481–488, 2015.
- [KZ03] Maxim Kontsevich and Anton Zorich. Connected components of the moduli spaces of Abelian differentials with prescribed singularities. *Inventiones Mathematicae*, 153(3):631–678, 2003.
- [Mas82] Howard Masur. Interval exchange transformations and measured foliations. *Annals of Mathematics. Second Series*, 115(1):169–200, 1982.
- [Mas88] Howard Masur. Lower bounds for the number of saddle connections and closed trajectories of a quadratic differential. Holomorphic functions and moduli I, Proc. Workshop, Berkeley/Calif. 1986, Publ., Math. Sci. Res. Inst. 10, 215–228., 1988.
- [Mas90] Howard Masur. The growth rate of trajectories of a quadratic differential. *Ergodic Theory and Dynamical Systems*, 10(1):151–176, 1990.
- [Mir13] Maryam Mirzakhani. Growth of Weil-Petersson volumes and random hyperbolic surface of large genus. *Journal of Differential Geometry*, 94(2):267–300, 2013.
- [MP19] Mariam Mirzakhani and Bram Petri. Lengths of closed geodesics on random surfaces of large genus. *Commentarii Mathematici Helvetici*, 94(4):869–889, 2019.
- [MRR22] Howard Masur, Kasra Rafi, and Anja Randecker. Expected covering radius of a translation surface. *International Mathematics Research Notices*, 2022(10):7967–8002, 2022.
- [MS91] Howard Masur and John Smillie. Hausdorff dimension of sets of nonergodic measured foliations. *Annals of Mathematics. Second Series*, 134(3):455–543, 1991.
- [Sau18] Adrien Sauvaget. Volumes and Siegel-Veech constants of $\mathcal{H}(2G - 2)$ and Hodge integrals. *Geometric and Functional Analysis (GAFA)*, 28(6):1756–1779, 2018.
- [Val] Hunter Vallejos.
- [Vee82] William A. Veech. Gauss measures for transformations on the space of interval exchange maps. *Annals of Mathematics. Second Series*, 115(1):201–242, 1982.
- [Vee98] William A. Veech. Siegel measures. *Annals of Mathematics. Second Series*, 148(3):895–944, 1998.

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF CHICAGO, CHICAGO, IL.

Email address: `masur@math.uchicago.edu`

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF TORONTO, TORONTO, ON

Email address: `rafi@math.toronto.edu`

INSTITUTE FOR MATHEMATICS, HEIDELBERG UNIVERSITY, HEIDELBERG

DEPARTMENT OF MATHEMATICS, SAARLAND UNIVERSITY, SAARBRÜCKEN

Email address: `randecker@mathi.uni-heidelberg.de`