The horocycle flow on the moduli space of translation surfaces

Jon Chaika and Barak Weiss

Abstract

We survey some results on the dynamics of the horocycle flow on the moduli space of translation surfaces. We outline proofs of some recent results, obtained by the authors in collaboration with John Smillie, and pose some open questions.

INTERNATIONAL CONGRESS OF MATHEMATICIANS 2022 JULY 6–14 SAINT PETERSBURG

Published by EMS Press. DOI 10.4171/ICM2022/?
Proc. Int. Cong. Math. 2022, Vol. ?, pp. 2–20

1. Introduction

The study of dynamics on moduli spaces of translation surfaces has been undergoing intensive growth over the last two decades. This subdomain of ergodic theory lies at the crossroads of dynamics of Lie group actions and geometry of surfaces and has close connections with the theory of rational billiards, interval exchange transformations, Teichmüller theory, algebraic geometry, number theory, mathematical physics, and more. The foundations of the theory were laid down by Masur and Veech in the 1980's, in work motivated by a conjecture of Keane about interval exchange transformations. Through the efforts of many mathematicians (see the ICM proceedings contributions [11, 16, 23, 30, 49], and the survey [33] about billiards in this volume) we now know a great deal about the dynamics on these spaces. As entry points we recommend the surveys [17, 24, 26, 44, 46–48].

Our focus in this survey will be results and open questions concerning the dynamics of the horocycle flow. Much of the work on dynamics on spaces of translation surfaces has been motivated by a fruitful analogy with the study of Lie group actions on homogeneous spaces. In such a putative dictionary, the horocycle flow on moduli spaces corresponds to a unipotent flow on a homogeneous space, for which Ratner [31] famously showed that all orbit-closures and invariant measures admit a nice algebraic description. The celebrated 'magic wand' theorems of Eskin, Mirzakhani and Mohammadi [14, 15], which we will discuss briefly below, may be regarded as providing positive evidence for the existence of a corresponding picture for moduli spaces of translation surfaces. However, as we will see, the emerging picture for the horocycle flow in moduli spaces is more complicated than this simple analogy might suggest.

2. Definitions and background

There are several alternative points of view concerning the definitions of translation surfaces, their moduli spaces, and the $SL_2(\mathbb{R})$-action on them, see Definitions 1, 4, and 5 of [24]. See the surveys mentioned above for more information, and alternative definitions, and see [4, §2] for a more detailed treatment following the point of view we will take here.

A *polygonal surface* (which we will also call a *polygonal presentation of a translation surface*) is a finite collection of polygons in the plane, equipped with a partition of the sides into pairs of parallel sides of equal length and opposite orientation, which we identify by translations.

If e, e' is a pair of identified sides, then there is a unique translation $\varphi = \varphi_{e,e'}$ with $\varphi(e) = e'$, and we say that each $x \in e$ is identified with $\varphi(x) \in e'$. The identifying maps $\{\varphi_{e,e'}\}$ generate an equivalence relation on the polygonal surface. For points in the interior of polygons, the equivalence class is a singleton; for points in the interior of a side, it is a pair of points; and for vertices, it is some finite set of vertices. The union of polygons has a topology as a subset of Euclidean space, and we endow the polygonal surface with the quotient topology for the equivalence relation just defined. Thus the polygonal surface becomes a compact oriented surface. We make the further requirement that it is connected.

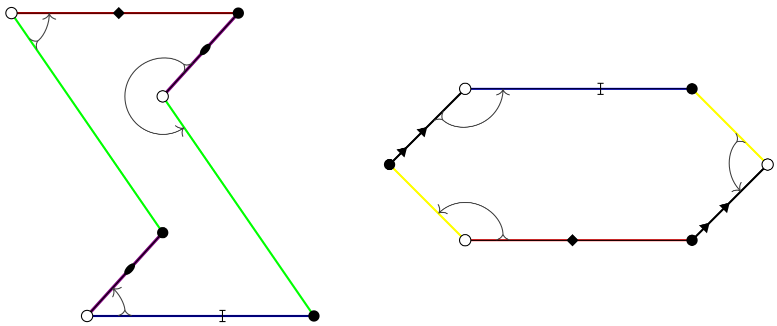

Figure 1 A polygonal surface. Parallel edges (in case of ambiguity, the ones with the same marking) are identified by translations, and the points marked with • and ∘ represent two singularities, each of order 1. The rotating arc around singularity ∘ measures its turning angle of 4π.

A polygonal surface inherits some geometric structures from the plane. Each point has a cone angle which measures the total turning angle made by a curve around the point. At points which are interior points of polygons or of edges, the turning angle is 2π, and for vertices of polygons, it is $2\pi(1+k)$ for some integer $k \geq 0$ measuring the *excess in angle*. Points for which the excess in angle is positive are called *singularities*, and the excess in angle of a singularity is its *order*. One defines the *area* of a polygonal surface, as the sum of the areas of the polygons. The surface also inherits the notion of a *straightline flow* in any direction. This is defined by extending the motion along a straight line by applying the maps $\varphi_{e,e'}$. If a straightline flow reaches a singularity, the straightline trajectory does not extend past the singularity, and thus the straightline flow in a given direction is defined for all times, only on a dense G_δ subset of the polygonal surface. A finite straightline flow trajectory which begins and ends at singular points is called a *saddle connection*. One can also measure the total horizontal and vertical displacement along an oriented path α in a polygonal surface M; i.e. the total amount travelled in the horizontal and vertical directions, when travelling along the path. We denote by $\mathrm{hol}_M(\alpha) \in \mathbb{R}^2$ the *holonomy vector* whose components are these horizontal and vertical displacements. See Figure 2.

There is a *scissors congruence* equivalence relation on polygonal surfaces, generated by the following three operations:

(a) subdividing a polygon into two polygons by adding a diagonal (in this case the two new edges are‘both sides’ of the new diagonal and they are identified);

(b) the inverse operation of amalgamating two polygons separated by an edge into a larger one by deleting a diagonal; and

(c) translating polygons by translations.

See Figure 3.

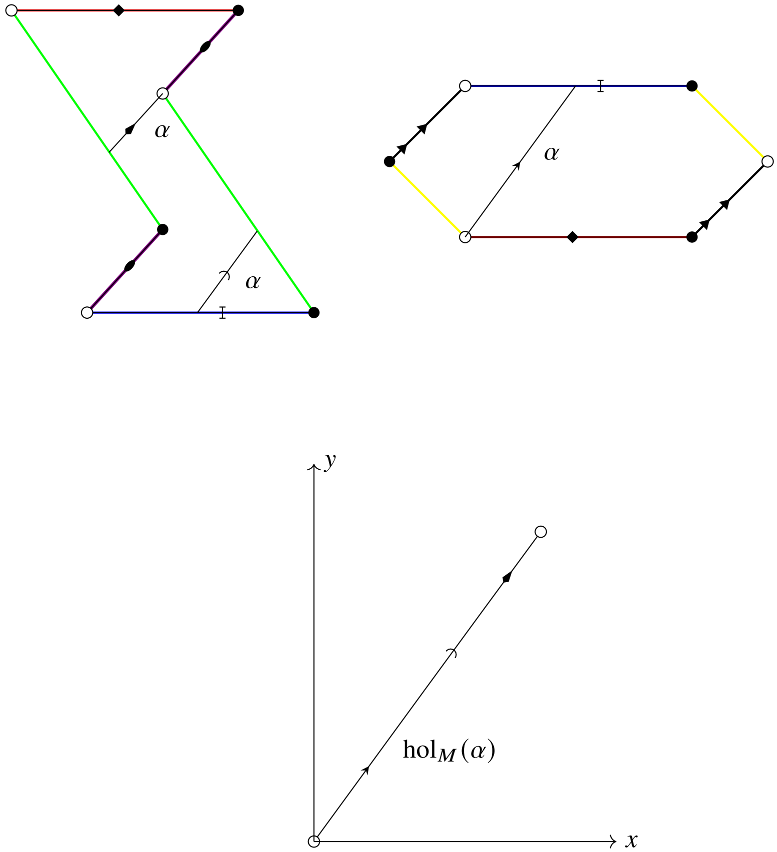

Figure 2 Measuring $\mathrm{hol}_M(\alpha)$ for a saddle connection α.

A scissors equivalence class of polygonal surfaces is called a *translation surface*. The number of singularities of a fixed order, as well as the area and the holonomy vectors of piecewise linear paths, are the same for polygonal surfaces that are the same up to scissors congruence, and thus make sense on translation surfaces.

The collection of all translation surfaces with a fixed number of singularities of given orders is called a *stratum*; we denote by $\mathcal{H}(a_1, \ldots, a_r)$, where $a_1, \ldots, a_r$ are positive integers, the stratum of translation surfaces with r singularities, of orders $a_1, \ldots, a_r$. The group

$$G = \mathrm{SL}_2(\mathbb{R}) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} : ad - bc = 1 \right\}$$

acts on the plane by linear transformations. This action extends to an action on polygonal surfaces (by applying the same linear transformation to each polygon) and preserves scissors

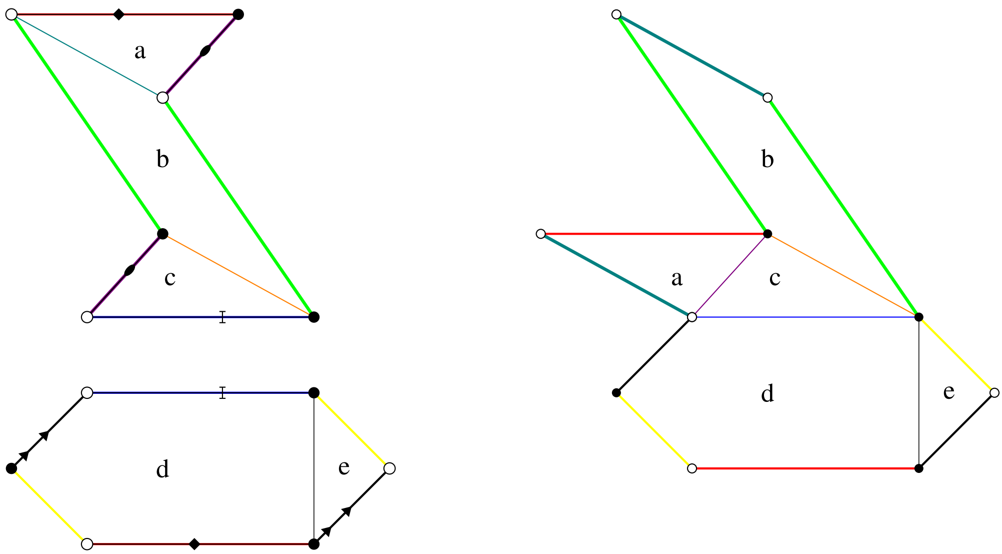

Figure 3 Two scissors equivalent polygonal surfaces.

equivalence, and thus acts on each stratum. The restriction of this action to the group

$$\{g_t : t \in \mathbb{R}\}, \quad \text{where } g_t = \begin{pmatrix} e^t & 0 \\ 0 & e^{-t} \end{pmatrix}$$

is called the *geodesic flow*. This action has been extensively studied, since the projections of its orbits to the moduli space of Riemann surfaces are parameterized geodesic paths with respect to the Teichmüller metric, and also since it provides a renormalization framework used for studying billiards and interval exchange transformations. Our focus will be on the *horocycle flow*, which is defined as the restriction of the G-action to the subgroup

$$U = \{u_s : s \in \mathbb{R}\}, \quad \text{where } u_s = \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix}. \tag{2.1}$$

We will pay special attention to orbit-closures for this action. For this we need to define a topology on a stratum $\mathcal{H}$. The topology we will use is a metrizable topology, characterized by the property that a sequence $\{M_j\}$ of translation surfaces converges to M as $j \to \infty$ if one can choose polygonal surfaces which are representatives for M and each of the M_j, such that for all large enough j, the polygonal surfaces have the same combinatorics (that is, the same number of polygons with the same number of sides and the same side identifications) and the vertices of the polygons comprising M_j converge to the corresponding vertices of the polygons comprising M. See [4, §2.4] for more details.

2.1. Some foundational results

The following fundamental results were proved in the 1980's and 1990's.

- The G-action preserves the area of a translation surface, and we denote by $\mathcal{H}_1(a_1, \ldots, a_r)$ the collection of area-one surfaces in $\mathcal{H}(a_1, \ldots, a_r)$. There is a natural smooth G-invariant measure of full support on the spaces $\mathcal{H}_1(a_1, \ldots, a_r)$, derived from the Lebesgue measure in 'period coordinates' via a 'cone construction', which was constructed by Masur[20] and Veech [40], and is now referred to as the *Masur-Veech measure*. Masur, Veech, and Masur-Smillie [25] proved that it is finite. Kontsevich and Zorich [19] classified the connected components of strata.

- Masur [20] and Veech [38] showed interval exchange transformations can be suspended and understood via straighline flows on translation surfaces, and that the $\{g_t\}$-action can be used to renormalize the straightline flow dynamics. They used this approach to settle a conjecture of Keane concerning unique ergodicity of interval exchange transformations.

- Masur [20] and Veech [39] showed that the G-action is ergodic with respect to the Masur-Veech measure. This implies that there are dense orbits for the horocycle flow and the geodesic flow, and that these flows are both mixing.

- Veech [37] gave examples of $\mathbb{Z}/2\mathbb{Z}$ skew products of rotations, that can be interpreted as flows on translation surfaces [26]. In these translation surfaces one has directions in which the straightline flow is minimal but not uniquely ergodic. Similar examples were also independently constructed by Sataev [32]. This phenomenon of minimality without unique ergodicity of the straightline flow will play an important role in our discussion. Masur [22] established a link between nondivergence of geodesic trajectories and unique ergodicity of foliations. Masur and Smillie [25] showed that while it is rare, there are abundant examples of minimal and not uniquely ergodic flows on translation surfaces.

- Veech [41] gave example of surfaces whose G-orbit carries a finite G-invariant measure (such surfaces are now known as *Veech surfaces*). Using the connection to the G-dynamics, he showed that for Veech surfaces, the straightline flow dynamics admits a complete description.

- Masur [21] showed that G-orbits are never bounded, and used this to show the existence of periodic trajectories for rational billiards. On the other hand, Smillie (see [36,42]) showed that a G-orbit Gq is closed if and only if q is a Veech surface.

2.2. The analogy with Ratner's work, and the magic wand theorem

The results mentioned in §2.1 can be seen as counterparts of similar results in the setting of homogeneous flows. Around the end of the 20th century, several researchers began to speculate that there might be a translation surface analogue of Ratner's celebrated theorems on the action of groups generated by unipotents, acting on homogeneous spaces.

In particular, see [1,11], who noted the usefulness of obtaining analogues of Ratner's theorem for applications in geometry and dynamics of translation surfaces.

What makes Ratner's results so powerful are that they are able to shed light on the behavior of *every* orbit (in contrast to softer results in ergodic theory which describe the behavior of typical orbits). Indeed, this analogy led to the hope that it might be possible to completely classify all invariant measures and all orbit-closures for the G-action and the U-action, as these are the only two connected subgroups of G (up to conjugation) that are generated by unipotent one parameter subgroups. McMullen [27] established such a result for the G-action in genus two, see also Calta [6] for earlier strong results in this direction. These results gave further impetus to work in this direction. The search was officially on when Zorich published an influential survey [48] with a section titled 'hope for a magic wand'. For the G-action, the conjecture was confirmed in spectacular fashion by Eskin, Mirzakhani and Mohammadi in [14, 15]. This work has revolutionized the study of dynamics on translation surfaces and has already had many applications in geometry which we do not survey here. As a sample of their results, we have the following:

Theorem 2.1 (Eskin-Mirzakhani-Mohammadi, P-genericity). *For any translation surface q there is a measure ν whose support is the orbit-closure $\overline{Gq}$, and such that for any compactly supported continuous test function φ on the stratum containing q,*

$$\frac{1}{T}\int_0^T \int_0^1 \varphi(g_t u_s q)\,ds\,dt \underset{T\to\infty}{\longrightarrow} \int \varphi\, d\nu.$$

The measure ν is affine in natural coordinates, see [15, Def. 1.1] *for a precise statement.*

These developments left open the question of whether a similar result was possible for the U-action; i.e., is it possible to classify all the U-orbit closures in terms of some algebraic or geometric data? Can one understand all U-invariant ergodic measures, and the asymptotic distribution of averages along any U-orbit? While the focus of this survey is on the horocycle dynamics as an interesting subject in its own right, we note that positive answers to these questions would have far-reaching consequences for some counting problems associated with billiards and flat surfaces. However, as we will see in this survey, the behavior of U-orbits in strata of translation surfaces can be quite different from the behavior of unipotent trajectories in homogeneous spaces.

3. Behavior of individual horocycle orbits

3.1. Some early results

Using ideas of Kerckhoff, Masur and Smillie [18], Veech [43] showed that there is no orbit of the horocycle flow that diverges in $\mathcal{H}$. That is, for any $q \in \mathcal{H}$, there is a compact $K \subset \mathcal{H}$ such that the set of visit times

$$\{s > 0 : u_s q \in K\}$$

is unbounded. A quantitative strengthening of this result was obtained by Minsky and Weiss [28]: for any $q \in \mathcal{H}$ and any $\varepsilon > 0$ there is a compact subset $K \subset \mathcal{H}$ such that

$$\liminf_{T\to\infty} \frac{1}{T} |\{s \in [0,T] : u_s q \in K\}| > 1 - \varepsilon.$$

These results are parallels of quantitative nondivergence results of Dani and Margulis for unipotent flows on homogeneous spaces (see [10]). With these results in hand, Smillie and Weiss [35] classified the minimal sets for the U-action on $\mathcal{H}$. They showed that for any $q \in \mathcal{H}$, the orbit-closure $\overline{Uq}$ contains a *minimal set*, i.e., a closed U-invariant subset containing no proper closed U-invariant subsets. Furthermore, $\overline{Uq}$ is minimal if and only if the straight-line flow in the horizontal direction on the underlying surface M_q is completely periodic. It follows that any orbit-closure for the U-action contains such a horizontally completely periodic surface.

In some rather special settings, it was possible to completely classify the U-invariant measures and orbit-closures. For Veech surfaces, this follows from results in homogeneous dynamics, as was observed in [13]. The first result of this kind in a non-homogeneous setting is due to Eskin, Marklof and Witte Morris [12], who studied surfaces which are branched covers of Veech surfaces. This work was later extended by Calta and Wortman [7] and Bainbridge, Smillie and Weiss [4]. In [4], a complete classification of U-invariant measures and orbit-closures is given within the eigenform loci in genus two. In these loci, which are 5-dimensional G-orbit-closures in $\mathcal{H}(1,1)$ arising in McMullen's genus two classification, we have a complete understanding of the possible orbit-closures and invariant measures. In these examples one can observe some phenomena not present for the G-dynamics, for instance, orbit-closures which are manifolds with non-empty boundary and an infinitely generated fundamental group. Nevertheless, these partial results were all consistent with a putative 'magic wand theorem for horocycles'.

3.2. Recent results

The situation changed in our work [9]. In this paper we proved the following results. We recall that if μ is a measure on $\mathcal{H}$ and $q \in \mathcal{H}$, we say that *q is generic for μ* if for any compactly supported continuous function f on $\mathcal{H}$ one has

$$\frac{1}{T}\int_0^T f(u_s q)\, ds \underset{T\to\infty}{\longrightarrow} \int_{\mathcal{H}} f\, d\mu. \tag{3.1}$$

Theorem 3.1. *Let $\mathcal{H} = \mathcal{H}(1,1)$ be the stratum of genus two surfaces with two singular points. Then:*

(1) *There is a surface $q \in \mathcal{H}$ and a G-invariant ergodic measure μ on $\mathcal{H}$, such that q is generic for μ but* $\mathrm{supp}(\mu) \subsetneq \overline{Uq}$.

(2) *There is a surface $q \in \mathcal{H}$ which is not generic for any measure.*

(3) *There is a surface $q \in \mathcal{H}$ whose orbit-closure is a fractal, in the sense that the Hausdorff dimension of $\overline{Uq}$ is not an integer.*

We make some remarks to put these results in context. The third item in Theorem 3.1 is perhaps the most striking but the first two items are also in stark contrast with a 'magic wand paradigm'. Note that in (1), the support of the measure limit measure μ is not the closure of the orbit – compare Theorem 2.1 or Ratner's work. Also in (2) we see that there is no analogue of Theorem 2.1 for the horocycle flow.

The orbit-closure we construct in (3) has an explicit description. The precise statement requires some technical preparation and will not be discussed here, see [9, Thm. 1.8].

The stratum $\mathcal{H}(1,1)$ is the simplest one in which we are able to exhibit a surface satisfying (3) but it is likely that our method can be extended to many other strata. However we are not able to establish (1) and (3) in the stratum $\mathcal{H}(2)$. Note however that (2) holds in $\mathcal{H}(2)$ by work of Chaika, Khalil and Smillie [8].

In the next sections we will explain some of the ideas of [9], focusing on the proofs of (1) and (2).

4. Tremors

The dynamical properties of a horocycle flow trajectory are intimately related to those of the horizontal straightline flow on the corresponding surfaces. Following [9] we will use the notation q to refer to a surface in a stratum, and M_q to refer to the underlying translation surfaces. Although they are formally identical, we will use the symbol q when the dynamical system we are considering is primarily the G-action on the stratum $\mathcal{H}$ containing q, and we will use $M = M_q$ when we are considering the dynamics of the horizontal straightline flow on the underlying translation surface. From now on by straightline flow we always mean the horizontal straightline flow, which we will denote by $\{\phi_t\}$ (the surface on which the flow takes place will be clear from the context).

Let ν be a $\{\phi_t\}$-invariant measure on M. The simplest example is the Lebesgue measure on the individual polygons. The second simplest example is the restriction of Lebesgue measure to a polygonal subsurface which is $\{\phi_t\}$-invariant; for example, in Figure 1, the restriction of Lebesgue measure to one of the two hexagons in the picture (note that these hexagons are separated from each other by two horizontal saddle connections, and thus each is $\{\phi_t\}$-invariant). Finally, a more interesting example referred to earlier, is the case when the straightline flow is minimal but not uniquely ergodic; in that case there will be two or more mutually singular $\{\phi_t\}$-invariant measures, all supported on the entire surface M. By a standard result (see e.g. [26]), if the straightline flow is not minimal then the surface contains a horizontally invariant polygonal subsurface, and thus this list exhausts all possible cases. Note that for almost every surface M, with respect to the measures discussed in §2.1, the only $\{\phi_t\}$-invariant measure (up to scaling) is Lebesgue measure.

Let σ be any non-horizontal segment in M. Such a segment is known as a *cross-section*. We consider σ as a piece of a trajectory for a (non-horizontal) straightline flow, thus parameterizing it by an interval, where we choose the *positive orientation* on σ so that $\mathrm{hol}_M(\sigma) = (x_\sigma, y_\sigma)$ satisfies $y_\sigma > 0$. From ν we can construct a *cross-section measure* on

σ, via the formula

$$\beta_{\sigma,\nu}(A) = \beta(A) = \lim_{\varepsilon\to 0+} \frac{1}{\varepsilon} \nu\left(\{\phi_t(a) : a \in A, t \in [0,\varepsilon]\}\right).$$

This classical construction defines a bijection between straightline flow invariant measures on M and measures on σ which are invariant under the first return map to σ along horizontal lines (see [2])[1]. If $\nu = \text{Leb}$ is Lebesgue measure on M, the cross-section measure on σ (viewed as an interval via its parameterization) is a multiple of one dimensional Lebesgue measure. The system of measures $\beta_\nu = \{\beta_{\sigma,\nu} : \sigma \text{ is a cross-section}\}$ is an example of a *transverse measure* (corresponding to ν).[2] The transverse measure corresponding to Leb will be called the *canonical transverse measure*. Similarly, if ν is the restriction of Lebesgue measure to a polygonal subsurface, the cross section measure on each σ is the restriction of Lebesgue measure to a finite collection of subintervals. If $\{\phi_t\}$ is minimal but not uniquely ergodic, the cross-section measures are fully supported measures which may be distinct from Lebesgue measure.

Let us now express the action of the horocycle element u_s (notation as in (2.1)). Recall that u_s acts on polygonal surfaces by tilting or *shearing* polygons. The computation

$$\text{hol}_{u_s M}(\sigma) = \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_\sigma \\ y_\sigma \end{pmatrix} = \begin{pmatrix} x_\sigma + s y_\sigma \\ y_\sigma \end{pmatrix} = \text{hol}_M(\sigma) + \begin{pmatrix} s y_\sigma \\ 0 \end{pmatrix}$$

shows that the amount of tilting of a side σ of a polygon is proportional to its measure, with respect to the canonical transverse measure. We can tweak this definition and replace each appearance of y_σ with $\beta(\sigma)$, where β is a transverse measure on M. This idea gives rise to the *tremor map*. It takes as input a surface q, a transverse measure β, and a 'time parameter' s, and produces a new translation surface $q' = \text{trem}_{s,\beta}(q)$, where $M' = M_{q'}$ is defined by assigning to each side σ of a polygonal presentation of M_q the holonomy

$$\text{hol}_{M'}(s) = \text{hol}_M(\sigma) + \begin{pmatrix} s\beta(\sigma) \\ 0 \end{pmatrix}.$$

A basic observation is that this definition makes sense. That is, there is a polygon presentation of M_q for which the adjusted segments in the above definition still give a polygonal presentation of a surface M', and moreover M' does not depend on a particular choice of a polygonal presentation. Furthermore[3], it can be shown that $\text{trem}_{s,\beta}(q)$ is defined for all values of s. On the other hand, the tremor map is not a flow in the sense that the choice of β depends on the initial translation surface M_q. For most choices of M_q, the only choice for

1 More precisely, for this correspondence we need σ to intersect every straightline trajectory; this always happens when the straightline flow is minimal but it will be convenient to relax this condition and define $\beta_{\sigma,\nu}(A)$ for any σ.

2 In [9] we use a more general definition of transverse measures, but the only transverse measures we will need in this survey arise from straightline flow invariant measures via this construction.

3 Recall that in this survey we discuss a more restrictive class of transverse measures. This assertion is false in the more general context considered in [9].

β is the canonical transverse measure, and in that case $\mathrm{trem}_{s,\beta}(q)$ is nothing but the horocycle image $u_s q$. However, for surfaces M_q for which there are non-canonical transverse measures, we get other tremor paths $\{\mathrm{trem}_{s,\beta}(q) : s \in \mathbb{R}\}$.

Sometimes it will be helpful to ignore the dependence of $\mathrm{trem}_{s,\beta}(q)$ on s, and we will write $\mathrm{trem}_\beta(q) = \mathrm{trem}_{1,\beta}(q)$. Note that the multiple of a transverse measure by a positive scalar is also a transverse measure, and we have the obvious identity

$$\mathrm{trem}_{s\beta}(q) = \mathrm{trem}_{s,\beta}(q). \tag{4.1}$$

It is sometimes helpful to work with signed measures, which turns the set of all 'signed transverse measures' into a real vector space. We call elements of this vector space *signed foliation cycles*. One can extend the definition of a tremor to the case in which β is a signed foliation cycle, and then one obtains identities like (4.1) for all $s \in \mathbb{R}$. In this more general setup, the set of transverse measures forms a convex cone C_q^+ in the vector space of signed foliation cocycles. See [9, §6] for more details.

A crucial fact for our analysis is the fact that surfaces which are obtained from one another by a tremor 'have the same horizontal foliation'. To make this precise, in [9, §5], we show the existence of a homeomorphism $\psi = \psi_\beta : M_q \to M_{q'}$, which is a topological conjugacy between the straightline flow on M_q and the straightline flow on $M_{q'}$, i.e.,

$$\forall t \in \mathbb{R},\ \phi'_t \circ \psi = \psi \circ \phi_t, \tag{4.2}$$

where ϕ_t, ϕ'_t denote respectively the straightline flows on M_q and $M_{q'}$. The pushforward map ψ_* induces a bijection between the straightline flow invariant measures on M_q and those on $M_{q'}$, and thus between the cones of transverse measures $C_q^+, C_{q'}^+$. In particular, this holds when β is canonical, i.e., when $q' \in Uq$. Thus if M_q is not horizontally uniquely ergodic, then the same holds for $M_{q'}$. Furthermore, we have the relation

$$\forall s \in \mathbb{R},\quad u_s \mathrm{trem}_\lambda(q) = \mathrm{trem}_\lambda(u_s q), \tag{4.3}$$

where we have used λ to denote both a transverse measure on M_q, and its image under ψ_*. Formally this is a commutation relation between the maps $q \mapsto \mathrm{trem}_\lambda(q)$ and $q \mapsto u_s(q)$. Note however that off of a set of measure zero, any tremor is just the horocycle flow and (4.3) is just the relation $u_{s_1} \circ u_{s_2} = u_{s_2} \circ u_{s_1}$.

5. Some ideas in the proof of Theorem 3.1

5.1. U-orbits of tremored surfaces almost track U-orbits

The starting point for our analysis is the following observation:

Proposition 5.1. *There is a proper complete metric* dist *on $\mathcal{H}$, inducing the topology, such that the following holds. Let $q \in \mathcal{H}$ such that M_q admits a non-canonical transverse measure $\beta = \beta_\nu$, where ν is a straightline flow invariant measure satisfying $\nu \ll$ Leb. Let $q' = \mathrm{trem}_\beta(q)$. Then*

$$\sup_{s \in \mathbb{R}} \mathrm{dist}(u_s q, u_s q') < \infty. \tag{5.1}$$

Note that by (4.3), $u_s q' = \mathrm{trem}_\beta(u_s q)$. A useful (but imprecise) heuristic explanation of (5.1) is that a fixed tremor can only move points a bounded distance. The metric dist appearing in Proposition 5.1 was introduced in [3].

A more detailed analysis of the function $s \mapsto \mathrm{dist}(u_s q, u_s q')$ appearing in (5.1) yields the following statement.

Theorem 5.2. *Let $\mathcal{M} = \overline{Gq} \subsetneq \mathcal{H}$ be a G orbit-closure, and let ν be the G-invariant ergodic measure on $\mathcal{M}$ (as in Theorem 2.1). Suppose q is generic for ν and M_q is horizontally minimal but not uniquely ergodic. Let β be a non-canonical transverse measure on M_q, such that $q' = \mathrm{trem}_\beta(q) \notin \mathcal{M}$. Then there is $s_0 \in \mathbb{R}$ so that the surface $q_0 = u_{s_0} q$ satisfies*

$$\forall \varepsilon > 0, \quad \frac{1}{T}|\{s \in [0,T] : \mathrm{dist}(u_s q', u_s q_0) \geq \varepsilon\}| \underset{T\to\infty}{\longrightarrow} 0. \tag{5.2}$$

Note that $q_0 \in \mathcal{M}$ is also generic for ν, since q is. If β were the canonical transverse measure then $q' = u_{s_0} q = q_0$ and (5.2) would be vacuously true. The result asserts that when $q' \notin \mathcal{M}$, the trajectory of q' nevertheless spends all but a negligible proportion of its time arbitrarily close to the trajectory of a generic point. In particular, it 'falls back on $\mathcal{M}$'. Since genericity is not affected by modifying a trajectory on a set of zero measure, we see that Theorem 5.2 implies (1) of Theorem 3.1, provided one can find examples of $\mathcal{M}$ and q for which the conditions of Theorem 5.2 are satisfied.

That such examples exist follows from the genericity results in [4]. Indeed, in the setting of eigenform loci in $\mathcal{H}(1,1)$ studied in that paper, the condition of having a minimal but not uniquely ergodic horizontal straightline flow does not have any effect on the asymptotic distribution of a horocycle orbit. In the simplest of these examples, $\mathcal{M}$ can be taken to be the collection of surfaces in $\mathcal{H}(1,1)$ which admit a 2:1 branched covering of a torus (this orbit-closure is denoted by $\mathcal{E}_4$ in McMullen's classification [27]).

We now explain the idea behind the proof of (5.2). It is useful to view a transverse measure as a cohomology class. Indeed, a transverse measure (or indeed, a signed foliation cocycle) assigns a real number to any positively oriented transverse segment on M_q. One can check that the assignment $\sigma \mapsto \beta_{\sigma,\nu}(\sigma)$ (where $\beta = \{\beta_{\sigma,\nu}\}$) is a cochain representing a cohomology class in $H^1(M_q, \Sigma_q; \mathbb{R})$, where Σ_q is the set of singularities. Consider the vector bundle $\mathcal{B}$ over $\mathcal{H}$ for which the fiber over q is $H^1(M_q, \Sigma; \mathbb{R})$, that is

$$\begin{array}{ccc} H^1(M_q, \Sigma_q; \mathbb{R}) & \longrightarrow & \mathcal{B} \\ & & \downarrow \\ & & \mathcal{H} \end{array}$$

The bundle $\mathcal{B}$ has a simple description as a sub-bundle of the tangent bundle of $\mathcal{H}$, with the pair (q, β) representing the tangent direction of the curve $s \mapsto \mathrm{trem}_{s,\beta}(q)$[4]. In particular, $\mathcal{B}$ has a natural topology, and in this topology the set of cones of transverse

4 To make this description precise one should work in the category of orbifold bundles.

measures C_q^+ is closed (see [9, §4.1, §13]). That is, if $\beta_n \in H^1(M_{q_n}, \Sigma_{q_n}; \mathbb{R})$ are cohomology classes represented by transverse measures, and $(q_n, \beta_n) \to (q_\infty, \beta_\infty)$ as elements of $\mathcal{B}$, then β_∞ is also represented by a transverse measure on M_{q_∞}. Furthermore, the map

$$\mathcal{B} \to \mathcal{H}, \quad (q, \beta) \mapsto \mathrm{trem}_\beta(q) \tag{5.3}$$

is continuous with respect to the topology on $\mathcal{B}$.

This basic fact seems to contradict our previous heuristic that surfaces which are in the same U-orbit have the same cone of transverse measures. For suppose $q_n = u_n q$ for $u_n \in U$, and q and thus all of the q_n have a non-canonical transverse measure, but $q_n \to q_\infty$ where q_∞ has a uniquely ergodic straightline flow. Then we have that all the q_n have the same fixed cone C_q^+ of transverse measures, containing both the canonical transverse measure and a non-canonical one, while at the same time this cone of transverse measures converges to $C_{q_\infty}^+$, which is just the ray generated by the canonical transverse measure. How is this possible?

The answer is that, with respect to any reasonable metric on $\mathcal{B}$, the bijection sending C_q^+ to $C_{q_n}^+$ is far from being an isometry. One can define norms $\|\cdot\|_q$ on each $H^1(M_q, \Sigma_q; \mathbb{R})$, which are continuous in the bundle topology, with respect to which the unit length transverse measures $\{\beta \in C_{q_n}^+ : \|\beta\|_{q_n} = 1\}$ all converge to the unique unit length transverse measure in $C_{q_\infty}^+$. Thus, the cones $C_{q_n}^+$, although of dimension > 1 and in bijection with each other, 'collapse' down to the ray $C_{q_\infty}^+$.

Using this idea, in the proof of Theorem 5.2 we show that for any $\varepsilon > 0$, there is an open set $\mathcal{U} \subset \mathcal{M}$ containing all the uniquely ergodic surfaces, such that for any $q_1 \in \mathcal{U}$, the diameter of $\{\beta \in C_{q_1}^+ : \|\beta\|_{q_1} = 1\}$ is at most ε. By genericity, the orbit Uq spends all but a negligible proportion of its time in $\mathcal{U}$, and since the map (5.3) is continuous for the metric dist, (5.2) follows.

5.2. From genericity to lack of genericity

Recall that from Birkhoff's theorem, any ergodic U-invariant measure assigns full measure to its generic points. It is sometimes useful to work with *quasi-generic points* instead. These are defined as points q for which there is a sequence $T_n \to \infty$ such that for all compactly supported continuous test functions f,

$$\frac{1}{T_n}\int_0^{T_n} f(u_s q)\, ds \underset{n\to\infty}{\longrightarrow} \int f\, d\mu.$$

Note that a generic point is quasi-generic, but a point can be quasi-generic for two measures. If q_0 is quasi-generic for two measures μ and ν, we can take f for which $\int f\, d\mu \neq \int f\, d\nu$ to see that the limit $\lim_{T\to\infty} \frac{1}{T}\int_0^T f(u_s q_0)\, ds$ does not exist, and in particular, q_0 is not generic for any measure. Thus for a given dynamical system, the condition that there are distinct invariant measures, and every point is generic for one of them, implies that there are no points which are quasi-generic for two different measures. Recall that this condition is satisfied for unipotent flows on homogeneous spaces (by Ratner's work), as well as for some averages on moduli spaces of translation surfaces (e.g. the horocycle flow in the settings of [4,7,12], or two dimensional averages for G-invariant measures as in Theorem 2.1).

To see that this is quite restrictive, we note that the set of quasi-generic points for some measure ν is a G_δ set. Indeed, let $\{f_j\}_{j\in\mathbb{N}}$ be a dense countable collection of continuous compactly supported test functions. For each $j \in \mathbb{N}$, each $T > 0$, and each $\varepsilon > 0$, continuity of the action implies that

$$\mathcal{U}_{j,T,\varepsilon} = \left\{ q \in \mathcal{H} : \text{ for } i = 1, \dots, j, \ \left| \frac{1}{T} \int_0^T f_i(u_s q)\, ds - \int f_i\, d\mu \right| < \varepsilon \right\}$$

is open. The set of quasi-generic points can be written as

$$\bigcap_{j=1}^{\infty} \bigcap_{k=1}^{\infty} \bigcup_{T=k}^{\infty} \mathcal{U}_{j,T,1/j},$$

proving the claim.

By the Baire category theorem, any two dense G_δ subsets intersect. Let $\mathcal{M}$ be as in Theorem 5.2, let μ and ν be the fully supported G-invariant measures on $\mathcal{H}(1,1)$ and on $\mathcal{M}$ respectively. By Birkhoff's theorem, the set of generic points for μ is dense in $\mathcal{H}(1,1)$. Thus item (2) of Theorem 3.1 follows from Theorem 5.2 and the following:

Proposition 5.3. *The set of surfaces of the form*

$$\{\mathrm{trem}_\beta(q) : q \in \mathcal{M} \textit{ is generic for } \nu \textit{ and horizontally minimal}, \beta \in C_q^+\} \tag{5.4}$$

is dense in $\mathcal{H}(1,1)$.

In order to prove this statement we recall the observation 'if q' is obtained from q by a tremor then M_q and $M_{u_s q}$ have the same transverse measures', which we discussed above in connection with (4.3). We add to it the additional observation 'if q' is obtained from q by the $\{g_t\}$-action then M_q and $M_{q'}$ have the same transverse measures'. This is proved using a similar idea to the comparison homeomorphism of **[9, §5]**. Namely, the definition of the g_t action shows that if $q' = g_{t_0} q$ then there is a homeomorphism $M_q \to M_{q'}$ which intertwines the straightline flow up to a time change. That is, if $\{\phi_t\}$ and $\{\phi'_t\}$ denote respectively the horizontal straightline flows on M_q and $M_{q'}$, and $\psi : M_q \to M_{q'}$ is the map obtained by acting on a polygonal presentation as in the definition of the g_t action, then

$$\forall t \in \mathbb{R}, \ \phi'_{e^{t_0} t} \circ \psi = \psi \circ \phi_t.$$

It follows from this that analogously to (4.3), one has

$$\forall \beta \in C_q^+, \ t \in \mathbb{R}, \ g_t \mathrm{trem}_{s,\beta}(q) = \mathrm{trem}_{e^t s,\beta}(g_t q)$$

(where we consider β simultaneously as belonging to C_q^+ and $C_{g_t q}^+$ via the above bijection). Together with (4.3), we find that the set of surfaces $\mathcal{F}$ defined in (5.4) is invariant under both flows $\{g_t\}$, $\{u_s\}$, and hence, by Theorem 2.1, $\overline{\mathcal{F}}$ is G-invariant. Moreover $\mathcal{M} \subsetneq \overline{\mathcal{F}}$, and examining the possibilities for $\mathcal{F}$ in McMullen's classification **[27]** gives that $\mathcal{F}$ is dense in $\mathcal{H}(1,1)$.

6. Questions

There are many open questions about the horocycle flow on strata of translation surfaces. We list some of them. We begin with one of the most outstanding questions in the field:

Question 1. *Is there an 'exotic' U-ergodic measure? For example, measures whose support has non-integer Hausdorff dimension, or fully supported measures that differ from Masur-Veech measure.*

A precise statement of the above question is tricky, because Calta [6] and Smillie and Weiss [34] gave examples of U-invariant ergodic measures whose support is a manifold with boundary and infinitely generated fundamental group.

The following question is motivated by renormalization dynamics:

Question 2. *If ν is a U-invariant ergodic measure that is not G-invariant, are there two G-invariant ergodic measures μ_-, μ_+ so that*

- $\mathrm{supp}(\mu_-) \subsetneq \mathrm{supp}(\mu_+)$
- $g_t \nu \underset{t \to +\infty}{\longrightarrow} \mu_+$
- $g_t \nu \underset{t \to -\infty}{\longrightarrow} \mu_-$.

Note that we consider the zero measure to be a G-invariant ergodic measure. Even the special case of measures supported on periodic horocycles (where μ_- is the zero measure) is open and very interesting. The same question is also interesting for horocycle orbit closures.

Some basic questions on the topological dynamics of the horocycle flow are open. To us, the following is the most outstanding example.

Question 3. *Is the horocycle flow recurrent as a topological dynamical system? That is, is it true that for every $q \in \mathcal{H}$ there exists a sequence $t_i \nearrow \infty$ so that $u_{t_i} q \underset{i \to \infty}{\longrightarrow} q$?*

In the realm of orbit closures:

Question 4. *In [9] we construct an exotic U orbit-closure, which is the orbit closure of the tremor of a translation surface in an 'eigenform locus.' What are all the orbit closures of tremors of translation surfaces in eigenform loci that do not have horizontal saddle connections? Do they all have the description given in [9, Eq. (1.8)]? Informally, is any such orbit-closure the set of all surfaces obtained from tremoring surfaces in a given eigenform locus by at most a certain fixed amount?*

Of course we are also interested in other horocycle orbit-closures, including those that arise from tremors of surfaces in proper G-orbit closures outside of $\mathcal{H}(1,1)$.

In special cases (see [4,7,12,13]) the horocycle flow has been shown to behave much like it does in homogeneous settings. For example, every point is generic for some U-invariant ergodic measure. These examples are all *rank-one loci* in the sense of [45]. This motivates the following general question.

Question 5. *In the special setting of rank-one loci, what can be said about the behavior of the hororocycle flow?*

There is a growing dictionary between the earthquake flow and the horocycle flow. This dictionary was initiated by Mirzakhani [29], who used it to prove that the earthquake flow is ergodic. Calderon and Farre [5] have added to this dictionary and extended it to other actions, which has allowed them to showcase additional behavior of the earthquake flow. It is interesting to see whether some of these results can be proven directly in the setting of earthquake flows and if any arguments in the setting of earthquake flows can be used to show new behavior of horocycle flows on strata.

Acknowledgments

We are grateful to John Smillie, who is our collaborator and co-author for much of the work described in this survey. We are also grateful to CIRM, where this collaboration began.

Funding

J. Chaika was supported in part by a Poincare chair, a Warnock chair, NSF grants DMS-1452762 and DMS-2055354. B. Weiss was supported in part by BSF grant 2016256 and ISF grant 2919/19.

References

[1]

[2] W. Ambrose, Representation of ergodic flows. *Ann. of Math. (2)* **42** (1941), 723–739

[3] A. Avila, S. Gouëzel, and J.-C. Yoccoz, Exponential mixing for the Teichmüller flow. *Publ. Math. Inst. Hautes Études Sci.* (2006), no. 104, 143–211

[4] M. Bainbridge, J. Smillie, and B. Weiss, Horocycle dynamics: new invariants and the eigenform loci in the stratum $\mathcal{H}(1, 1)$, preprint, to appear in Mem. AMS, available on arXiv at arxiv.org/abs/1603.00808

[5] A. Calderon and J. Farre, Shear-shaped cocycles for measured laminations and ergodic theory of the earthquake flow, preprint, available on arXiv at arxiv.org/abs/2102.13124

[6] K. Calta, Veech surfaces and complete periodicity in genus two. *J. Amer. Math. Soc.* **17** (2004), no. 4, 871–908

[7] K. Calta and K. Wortman, On unipotent flows in $\mathscr{H}(1, 1)$. *Ergodic Theory Dynam. Systems* **30** (2010), no. 2, 379–398

[8] J. Chaika, O. Khalil, and J. Smillie, On the space of ergodic measures for the horocycle flow on strata of abelian differentials, preprint, available on arXiv at https://arxiv.org/abs/2104.00554

[9] J. Chaika, J. Smillie, and B. Weiss, Tremors and horocycle dynamics on the moduli space of translation surfaces, preprint, available on arXiv at arxiv.org/abs/2004.04027

[10] S. G. Dani, On invariant measures, minimal sets and a lemma of Margulis. *Invent. Math.* **51** (1979), no. 3, 239–260

[11] A. Eskin, Counting problems and semisimple groups. In *Proceedings of the International Congress of Mathematicians, Vol. II (Berlin, 1998)*, pp. 539–552, Extra Vol. II, 1998

[12] A. Eskin, J. Marklof, and D. Witte Morris, Unipotent flows on the space of branched covers of Veech surfaces. *Ergodic Theory Dynam. Systems* **26** (2006), no. 1, 129–162

[13] A. Eskin, H. Masur, and M. Schmoll, Billiards in rectangles with barriers. *Duke Math. J.* **118** (2003), no. 3, 427–463

[14] A. Eskin and M. Mirzakhani, Invariant and stationary measures for the $\mathrm{SL}(2,\mathbb{R})$ action on moduli space. *Publ. Math. Inst. Hautes Études Sci.* **127** (2018), 95–324

[15] A. Eskin, M. Mirzakhani, and A. Mohammadi, Isolation, equidistribution, and orbit closures for the $\mathrm{SL}(2,\mathbb{R})$ action on moduli space. *Ann. of Math. (2)* **182** (2015), no. 2, 673–721

[16] G. Forni, Asymptotic behaviour of ergodic integrals of 'renormalizable' parabolic flows. In *Proceedings of the International Congress of Mathematicians, Vol. III (Beijing, 2002)*, pp. 317–326, Higher Ed. Press, Beijing, 2002

[17] G. Forni and C. Matheus, Introduction to Teichmüller theory and its applications to dynamics of interval exchange transformations, flows on surfaces and billiards. *J. Mod. Dyn.* **8** (2014), no. 3-4, 271–436

[18] S. Kerckhoff, H. Masur, and J. Smillie, Ergodicity of billiard flows and quadratic differentials. *Ann. of Math. (2)* **124** (1986), no. 2, 293–311

[19] M. Kontsevich and A. Zorich, Connected components of the moduli spaces of Abelian differentials with prescribed singularities. *Invent. Math.* **153** (2003), no. 3, 631–678

[20] H. Masur, Interval exchange transformations and measured foliations. *Ann. of Math. (2)* **115** (1982), no. 1, 169–200

[21] H. Masur, Closed trajectories for quadratic differentials with an application to billiards. *Duke Math. J.* **53** (1986), no. 2, 307–314

[22] H. Masur, Hausdorff dimension of the set of nonergodic foliations of a quadratic differential. *Duke Math. J.* **66** (1992), no. 3, 387–442

[23] H. Masur, Teichmüller space, dynamics, probability. In *Proceedings of the International Congress of Mathematicians, Vol. 1, 2 (Zürich, 1994)*, pp. 836–849, Birkhäuser, Basel, 1995

[24] H. Masur, Ergodic theory of translation surfaces. In *Handbook of dynamical systems. Vol. 1B*, pp. 527–547, Elsevier B. V., Amsterdam, 2006

[25] H. Masur and J. Smillie, Hausdorff dimension of sets of nonergodic measured foliations. *Ann. of Math. (2)* **134** (1991), no. 3, 455–543

[26] H. Masur and S. Tabachnikov, Rational billiards and flat structures. In *Handbook of dynamical systems, Vol. 1A*, pp. 1015–1089, North-Holland, Amsterdam, 2002
[27] C. T. McMullen, Dynamics of $\mathrm{SL}_2(\mathbb{R})$ over moduli space in genus two. *Ann. of Math. (2)* **165** (2007), no. 2, 397–456
[28] Y. Minsky and B. Weiss, Nondivergence of horocyclic flows on moduli space. *J. Reine Angew. Math.* **552** (2002), 131–177
[29] M. Mirzakhani, Ergodic theory of the earthquake flow. *Int. Math. Res. Not. IMRN* (2008), no. 3, Art. ID rnm116, 39
[30] M. Möller, Geometry of Teichmüller curves. In *Proceedings of the International Congress of Mathematicians—Rio de Janeiro 2018. Vol. III. Invited lectures*, pp. 2017–2034, World Sci. Publ., Hackensack, NJ, 2018
[31] M. Ratner, Interactions between ergodic theory, Lie groups, and number theory. In *Proceedings of the International Congress of Mathematicians, Vol. 1, 2 (Zürich, 1994)*, pp. 157–182, Birkhäuser, Basel, 1995
[32] E. A. Sataev, The number of invariant measures for flows on orientable surfaces. *Izv. Akad. Nauk SSSR Ser. Mat.* **39** (1975), no. 4, 860–878
[33] R. E. Schwartz, Survey lecture on billiards, to appear in the Proceedings of the ICM
[34] J. Smillie and B. Weiss, Examples of horocycle invariant measures on the moduli space of translation surfaces, in preparation
[35] J. Smillie and B. Weiss, Minimal sets for flows on moduli space. *Israel J. Math.* **142** (2004), 249–260
[36] J. Smillie and B. Weiss, Finiteness results for flat surfaces: a survey and problem list. In *Partially hyperbolic dynamics, laminations, and Teichmüller flow*, pp. 125–137, Fields Inst. Commun. 51, Amer. Math. Soc., Providence, RI, 2007
[37] W. A. Veech, A Kronecker-Weyl theorem modulo 2. *Proc. Nat. Acad. Sci. U.S.A.* **60** (1968), 1163–1164
[38] W. A. Veech, Gauss measures for transformations on the space of interval exchange maps. *Ann. of Math. (2)* **115** (1982), no. 1, 201–242
[39] W. A. Veech, The Teichmüller geodesic flow. *Ann. of Math. (2)* **124** (1986), no. 3, 441–530
[40] W. A. Veech, Moduli spaces of quadratic differentials. *J. Analyse Math.* **55** (1990), 117–171
[41] W. A. Veech, Erratum: "Teichmüller curves in moduli space, Eisenstein series and an application to triangular billiards". *Invent. Math.* **103** (1991), no. 2, 447
[42] W. A. Veech, Geometric realizations of hyperelliptic curves. In *Algorithms, fractals, and dynamics (Okayama/Kyoto, 1992)*, pp. 217–226, Plenum, New York, 1995
[43] W. A. Veech, Measures supported on the set of uniquely ergodic directions of an arbitrary holomorphic 1-form. *Ergodic Theory Dynam. Systems* **19** (1999), no. 4, 1093–1109

[44] Y. B. Vorobets, Plane structures and billiards in rational polygons: the Veech alternative. *Uspekhi Mat. Nauk* **51** (1996), no. 5(311), 3–42

[45] A. Wright, Cylinder deformations in orbit closures of translation surfaces. *Geom. Topol.* **19** (2015), no. 1, 413–438

[46] A. Wright, Translation surfaces and their orbit closures: an introduction for a broad audience. *EMS Surv. Math. Sci.* **2** (2015), no. 1, 63–108

[47] J.-C. Yoccoz, Interval exchange maps and translation surfaces. In *Homogeneous flows, moduli spaces and arithmetic*, pp. 1–69, Clay Math. Proc. 10, Amer. Math. Soc., Providence, RI, 2010

[48] A. Zorich, Flat surfaces. In *Frontiers in number theory, physics, and geometry. I*, pp. 437–583, Springer, Berlin, 2006

[49] A. Zorich, Geodesics on flat surfaces. In *International Congress of Mathematicians. Vol. III*, pp. 121–146, Eur. Math. Soc., Zürich, 2006

Jon Chaika
University of Utah, chaika@math.utah.edu

Barak Weiss
Tel Aviv University, barakw@tauex.tau.ac.il