IV. The Dynamics of Billiard Flows in Rational Polygons

Contents

Chapter 11 The Dynamics of Billiard Flows in Rational Polygons

J. Smillie

Billiard systems provide classic examples of simple mechanical systems. Among such systems the simplest are those that model the motion of a single particle in a region P of the plane. The trajectory of a particle in P is defined by requiring that the particle move in a straight line and at constant velocity in the interior of P and, when it hits the boundary of P at a point where the boundary is a smooth curve, the particle should reflect off of the boundary so that the angle of incidence is equal to the angle of reflection. If the trajectory hits a corner then there may be no good physical principal which selects one particular continuation. In this case the continuation of the trajectory is undefined.

In §2 we will explain how billiard trajectories are projections of orbits of a "geodesic flow" or "billiard flow" on a suitably defined tangent bundle. Thus the behavior of these trajectories is tied to the dynamics of the billiard flow.

Trajectories in planar billiard tables exhibit a wide range of behaviors. Two trajectories which are nearby but not parallel tend to diverge. Trajectories which are nearby and parallel remain parallel until they hit the boundary of P. Two features of the boundary may lead to divergence of nearby parallel trajectories: curvature and corners. The article by Bunimovich in this volume discusses the effects of curvature on planar billiard dynamics. If the sides of P are straight segments then curvature does not play a role. In this case nearby parallel trajectories can only diverge if they hit the boundary on opposite sides of a vertex. Regions with straight sides are called *polygonal billiard tables*. In this article we will consider the dynamics of billiard flows on polygonal tables.

The most familiar and best understood example of a planar polygonal billiard table is the square. The analysis of the dynamical properties of the billiard flow for the square dates to 1913 ([KS] see also [FK] and [HW]). The square has a number of properties that distinguish it from the general polygonal table. One key feature of the square is that each trajectory travels in only finitely many directions. This behavior is a characteristic feature of the class of rational billiards, where a polygonal billiard table with connected boundary is *rational* if each vertex angle has angular measure which is a rational multiple of π. In this article we will focus on the dynamics of billiard flows on rational polygons. For a discussion of general polygonal billiards see the survey articles by Tabachnikov [Ta] and Gutkin [Gu3].

There are several reasons for studying rational billiards. The dynamics of rational billiard tables are simple enough so that a theory can be developed yet they are sufficiently complex that many open questions remain. We will see in §6 that the study of rational billiard dynamics leads to the study of certain flows on moduli spaces which are themselves objects of great dynamical interest. Finally the fact that rational billiards are more complicated than "integrable" systems and yet not fully "chaotic" has led physicists to consider them as test cases for questions relating quantum dynamics to classical dynamics [cf BR]. While this leads to interesting question for future mathematical investigation, in this article we will deal exclusively with the "classical dynamics" of billiard tables.

The main questions that we will consider involve the distribution of billiard trajectories. Here are three specific questions.

(1) If we fix a table and fix a an initial direction of travel, what are the possible behaviors of the trajectories and how does this behavior depend on the initial position?
(2) How does the answer to question (1) change as we vary the direction?
(3) How does the answer to question (2) depend on the table?

§1. Two Examples

To give the flavor of rational billiard dynamics we will consider two examples, the square and the "divided rectangle", where the *divided rectangle* is the table obtained from the square by introducing a vertical reflective barrier in the center of the square which divides the square into two rectangles connected by an opening at the bottom (see figure). In fact the divided rectangle gives us a one parameter family of tables to consider, since we can adjust the length of the barrier. The billiard flow for the divided rectangle is easier to analyze than billiard flows on general rational tables, yet it exhibits many of the features seen in billiard flows on typical tables.

In the square, if two trajectories are parallel and start close together then they remain close together for all time, even if they hit the boundary on opposite sides of a vertex. This is an unusual feature for a rational polygon and accounts for some of the special properties of billiard trajectories in the square. It is no longer true in the divided rectangle, as we can see by considering two nearby parallel trajectories one of which hits the central barrier near its tip and one of which does not.

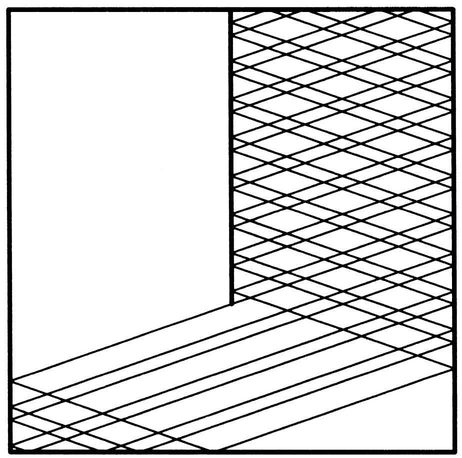

Fig. 1

In both of our examples the trajectories with rational slope have special properties. In the square, a trajectory with rational slope is periodic and the period is independent of the starting point. In the divided rectangle it is still true that all trajectories with rational slope are periodic but the period of the trajectory may depend on the initial point. In the square, each trajectory with irrational slope is dense. In contrast we can find trajectories in the divided rectangle with irrational slope which are not dense in the table (see Fig. 1). The first example of such behavior for polygonal tables was discovered by Galperin [Ga].

A dense orbit "fills up the table" but we can ask more specifically about the *rate* at which it fills up the table. We say that an orbit is *uniformly distributed* if the amount of time that it spends in a region is proportional to the area of the region. A trajectory which is uniformly distributed is necessarily dense. If the direction of the flow is irrational then every orbit in the square is not only dense but uniformly distributed. This is not the case for the divided rectangle; for certain barrier lengths there are orbits with irrational slope which are dense in the table but spend more time on the left of the barrier than on the right. It is not easy to illustrate this behavior with a computer picture but Fig. 2, which shows an ϵ-dense orbit which spends more time in the left half of the table than the right, is meant to be suggestive. At the end of §2 we will explain how the existence of dense but not uniformly distributed trajectories follows from an early theorem of Veech.

We say that a polygonal table has the *dichotomy property* if all non-singular orbits are closed or are uniformly distributed. The divided rectangle does not have this property (at least for most lengths of the gap) and there are reasons to think that this property is quite rare. Nevertheless we will see in §7 that

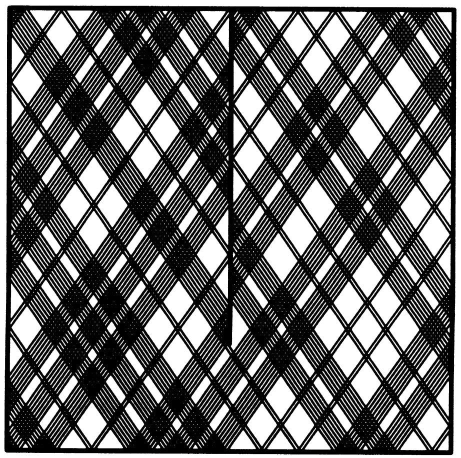

Fig. 2

there are in fact some interesting examples other than the square where the dichotomy property holds.

§2. Formal Properties of the Billiard Flow

In this section we will relate billiard trajectories to orbits of a billiard flow on an appropriately defined tangent bundle. We will investigate an invariant foliation on this tangent bundle and describe a technique from [ZK] which allows us to reduce questions about billiards to questions about the geodesic flow on certain singular surfaces.

We begin with some basic observations about billiard flows. Let P be a polygon in the plane which, for the moment, we do not assume to be rational. We will describe the construction of a "geodesic flow" or "billiard flow" on the unit tangent bundle to P whose orbits project to billiard trajectories on P which travel at unit speed. Let S^1 be the unit circle in $\mathbf{R}^2$ and let $T(\mathbf{R}^2) = \mathbf{R}^2 \times S^1$ be the unit tangent bundle of $\mathbf{R}^2$. The geodesic flow on $T(\mathbf{R}^2)$ induces a partially defined flow on $P \times S^1 \subset T(\mathbf{R}^2)$ where orbits fail to have continuations when they hit the boundary of P. We would like trajectories to reflect off the boundary, and the simplest way to achieve this is to identify certain inward and outward pointing vectors at points of the boundary of P. If e_i is an edge of P and $\rho_i : S^1 \to S^1$ represents the reflection through e_i then we identify (p, v) with $(p, \rho_i(v))$ for each $p \in e_i$. We define the tangent bundle of P, $T(P)$, to be $P \times S^1/\sim$ where $\sim$ is the equivalence relation generated by identifying (p, v) with $(p, \rho_i(v))$ as above. At the vertex p where the edge e_i meets the edge e_j we identify (p, v) with $(p, \gamma(v))$ for all γ in the group generated by ρ_i and ρ_j. Away from the vertices we can define a billiard flow on $T(P)$ whose trajectories project to billiard trajectories on P. This flow is continuous where it is defined.

The geodesic flow on $\mathbf{R}^2$ has a number of special properties, some of which are reflected in properties of billiard flows. Two tangent vectors (p, v) and (p', v') in $T(\mathbf{R}^2)$ are *parallel* if $v = v'$. The relation of being parallel gives an equivalence relation on $T(\mathbf{R}^2)$ which is preserved by the geodesic flow. The equivalence classes of this relation are copies of $\mathbf{R}^2$ which we can think of as leaves of a "parallel foliation" of $T(\mathbf{R}^2)$. Thus the geodesic flow on $\mathbf{R}^2$ can be decomposed as a family of "directional flows", one for each direction $v \in S^1$. We can view each of these directional flows as a flow on $\mathbf{R}^2$.

There is a related foliation of $T(P)$ which has corresponding properties. The parallel foliation of $T(\mathbf{R}^2)$ induces a foliation of $P \times S^1$. When two parallel billiard trajectories reflect off of the same edge they remain parallel. This implies that the identifications used to create the tangent bundle $T(P)$ in fact preserve the leaves of the parallel foliation. Thus there is a natural induced "parallel" foliation of $T(P)$ where leaves of this new foliation are obtained by gluing together the leaves of the parallel foliation of $P \times S^1$. Two points (p, v) and (p', v') are in the same leaf of this new foliation if there is some sequence of edges $e_{i_1}, \ldots, e_{i_n}$ so that $v' = \rho_{i_1} \circ \ldots \circ \rho_{i_n}(v)$. Let $\Gamma \subset O(2)$ be the group generated by the reflections in sides. The leaves of the foliation correspond to points in the orbit space S^1/Γ.

The properties of the parallel foliation of $T(P)$ depend on the cardinality of Γ. When P is a rational table then Γ is finite. In this case the "leaf space" S^1/Γ is an interval and the leaves of the foliation are closed surfaces. We will assume from now on that the polygon P is rational. Denote the interval S^1/Γ by I. For each $\theta \in S^1$ let M_θ be the surface corresponding to θ. Since the surfaces M_θ are invariant, the billiard flow on $T(P)$ decomposes into a family of *directional flows* on the surfaces M_θ.

The surfaces M_θ are constructed from copies of P which are glued along their edges by isometries. Since each surface M_θ is constructed by gluing together copies of P according to the same pattern all such surfaces can be identified with a single surface $\tilde{P}$. (If θ corresponds to an endpoint of I this is not quite true but this is a minor point which we will ignore.) The surface $\tilde{P}$ appears in [FK] in the case of the square. The general case was considered in [ZK] see also [RB].

Since the surface $\tilde{P}$ is built by gluing together polygons by isometries it has a natural metric space structure (cf [KS]). At points of $\tilde{P}$ corresponding to interior points of P or to edges of P the surface is locally isometric to $\mathbf{R}^2$. In particular there is a natural notion of parallel translation along paths which do not run through the vertices. The behavior at vertices is more complicated. Consider the following situation. Let $p_1 \ldots p_m$ be vertices in polygons $P_1 \ldots P_m$. Let θ_j be the vertex angle at p_j. Glue these polygons together in a cyclic pattern so that all vertices p_j are identified with a single point p. We say that the resulting space has a "cone type singularity" at p and we define the *cone angle at* p to be $\theta = \sum \theta_j$. If the cone angle is equal to 2π then the resulting surface is locally isometric to $\mathbf{R}^2$ at p. We can think of such points as a "removable singularities". Removable singularities arise at points in $\tilde{P}$ corresponding to vertices in P with vertex angles of the form π/n. We call points of $\tilde{P}$ at which the cone angle is not equal to 2π *vertices* of $\tilde{P}$.

When we construct the surface M_θ we are gluing together a finite number of copies of P each with a vector field on it. We perform the gluing so that the vector fields match along the edges of the polygons (though they may not match at the vertices). Since we can identify M_θ with $\tilde{P}$ we can think of these vector fields as vector fields on $\tilde{P}$. The vector fields that arise in this manner are precisely the "parallel vector fields" on $\tilde{P}$. A *parallel* vector field has the property that the vectors at any two points are parallel translates of one another. Thus we can think of the directional flows as the family of

(partially defined) flows generated by the collection of parallel vector fields on the single surface $\tilde{P}$.

The fact that $\tilde{P}$ has a parallel vector field implies that the cone angles are multiples of 2π. The vector field can be extended to a point p if and only if the cone angle at p is equal to 2π which is to say that the singularity is removable.

Any two parallel vector fields on $\tilde{P}$ commute. The fact that the billiard flow leaves invariant a decomposition of $T(P)$ into surfaces and on each surface there is a pair of commuting vector fields is reminiscent of the properties of integrable flows on manifolds (cf Chapter 6 §1 of this volume). There is an important distinction between typical billiard flows on $\tilde{P}$ and integrable flows which is related to the existence vertices. In the case of integrable flows the invariant surface is a torus, that is to say a surface of genus one. In the rational billiard case the presence of vertices allows the possibility that the surface can have genus greater than one.

The Gauss-Bonnet theorem shows that the surface $\tilde{P}$ has genus one precisely when all the singular points are removable. In this case the analogy with integrable systems is complete and such polygonal billiard tables are called *integrable*. Richens and Berry [RB] have introduced the term *quasi-integrable* for the more typical case when the genus of $\tilde{P}$ is greater than one.

The list of integrable tables is short. The square is integrable as are rectangles. The only other integrable polygons are the triangle with angles $\pi/4$, $\pi/4$ and $\pi/2$, the equilateral triangle and the triangle with angles $\pi/6$, $\pi/3$ and $\pi/4$. When P is the divided rectangle then $\tilde{P}$ has two vertices each with cone angle 4π. This surface has genus 2 so this example is not integrable. It is nevertheless related to the torus as it can be viewed geometrically as a branched double cover of the torus.

The surfaces $\tilde{P}$ belong to an interesting class of geometric objects called translation surfaces that we will define using a characterization due to Veech. Say that we have a surface M with a specified finite set $\Sigma \subset M$. A *translation structure* on M is given by an atlas of charts in $M - \Sigma$ taking their values in $\mathbf{R}^2$ so that the change of coordinate functions are restrictions of translations of $\mathbf{R}^2$. This atlas of charts induces a Riemannian metric on $M - \Sigma$ and we impose the requirement that M is the metric completion of $M - \Sigma$ with respect to this Riemannian metric. In this case we will say that M is a surface with a translation structure or we will say that M is a *translation surface*. To see that $\tilde{P}$ has a translation structure we let Σ be the set of vertices and we choose a pair of perpendicular parallel vector fields on $\tilde{P}$. Take as a system of charts diffeomorphisms $\phi_j : U_j \to \mathbf{R}^2$ which take these parallel vector fields to the vector fields ∂_x and ∂_y in $\mathbf{R}^2$.

There is a geometric structure closely related to a translation structure where the change of coordinate functions are allowed to have the form $v \mapsto \pm v + c$. These structures are called *admissible $\mathcal{F}$ structures* by Veech or *half-*

integral translation structures in [GJ2] and they arise in the study of quadratic differentials [EG].

The surfaces $\tilde{P}$ provide examples of translation surfaces but there are many translation surfaces that do not arise from polygons via this construction. For any translation structure the geodesic flow decomposes into a family of "directional flows", one for each direction in the unit circle. When the translation structure does come from a rational billiard table this geodesic flow is the same as the billiard flow we have defined. Even when the translation structure does not arise from a billiard table this geodesic flow still constitutes an interesting dynamical systems. Our questions (1), (2) and (3) are still relevant. We can add two questions to our list:

(4) What behaviors are generic for translation structures?
(5) To what extent is the dynamical behavior of billiard tables like the dynamical behavior of translation structures?

§3. The Flow in a Fixed Direction

In this section we survey some results which describe the flow on a translation surface in a fixed direction. A useful technique in studying these flows is to consider the first return map to a transverse interval. Let I denote an interval in the surface M transverse to the flow. Let p be a point in I. The forward trajectory of p will either return to I or hit a vertex. The set of points in I for which forward trajectories hit vertices is finite. These points divide the interval I into subintervals $I_1 \dots I_k$. The restriction of the first return map to one of these intervals is an orientation preserving isometry. In particular the first return map is an interval exchange transformation. (See Chapter 4, Section 2 of this volume.)

Criteria for the minimality of interval exchange transformations lead to criteria for the minimality of directional flows. To describe one such criterion we will introduce some terminology. We will call a geodesic segment which starts at a vertex and ends at a vertex but contains no vertices in its interior a *vertex connection* or when no confusion will arise simply an *edge*. These are sometimes called *saddle connections* in the literature. A vertex connection has a well defined direction.

Theorem 3.1. [ZK], [BKM]. *If there are no vertex connections in direction v then the flow in direction v is minimal.*

If we are given a transverse interval I then there are two ways in which the directional flow could fail to be minimal. First there could be a set of trajectories which never hit I. In this case the union of these trajectories is a manifold with non-empty boundary and the boundary consists of vertex connections. Second it might be the case that all trajectories hit I but that the

first return map is not minimal. In this case the Keane criterion for minimality of interval exchange transformations (Ch. 4, Theorem 2.1) shows that there must be a vertex connection. The paper of Arnoux [A1] provides a good reference for the relations between flows on surfaces and interval exchange transformations.

The measure theoretic behavior of interval exchange transformations is more subtle than the topological behavior. Examples of Keane [K] and Keynes and Newton [KN] show that it is possible for an interval exchange transformation to be minimal but not ergodic. On the other hand there are some strong limitations:

Theorem 3.2. [Ka], [V2]. *There is a constant N depending only on the polygon so that for each minimal direction there are at most N ergodic invariant measures for the directional flow.*

If there is only one ergodic component for a directional flow then that flow is *uniquely ergodic*. When the flow is uniquely ergodic all non-singular orbits are uniformly distributed. Of course when the flow is uniquely ergodic it is ergodic with respect to the natural Lebesgue measure.

Criteria for the unique ergodicity of interval exchange transformations lead to criteria for the unique ergodicity of directional flows. The criteria for determining unique ergodicity are more complicated than those for determining minimality. One approach is through a kind of renormalization operator. Renormalization operators occur in many areas of dynamical systems. These operators are maps defined on spaces of dynamical systems. Typically they act by replacing a map by an iterate of the original map restricted to a smaller domain and rescaling the domain.

For an interval exchange transformation defined on an interval I the induced map on a subinterval $I_1 \subset I$ is again an interval exchange transformation so it is natural to consider renormalization operators on the space of interval exchange transformations. Rauzy induction provides one such method. Rauzy induction gives a map $\mathscr{R}$ from the space of interval exchanges to itself and a finite partition of this space. Let α denote an interval exchange transformation. By considering the partition element containing $\mathscr{R}^n(\alpha)$ we assign to α a symbol sequence. In the simplest case when there are only two intervals Rauzy induction corresponds to the continued fraction algorithm. In the general case one can give a criterion for unique ergodicity of α in terms of the symbol sequence of α. (See [Ke], [Ra] of [V2].)

Unique ergodicity is not the only property about the distribution of orbits of an interval exchange transformation that can be deduced from information about Rauzy expansion. If an interval exchange transformation is uniquely ergodic then for any continuous function f on I and any point p the Birkhoff sums along the orbit of p converge to the integral of f. Zorich gives connections between the Lyapunov exponents of the Rauzy induction map and the rate of convergence of Birkhoff sums. See [Z1], [Z2], [Z3] and the joint paper with Kontsevich [KZ].

As an example of these methods we will consider the interval exchange associated with the divided rectangle. Let us start by considering a degenerate case when the gap has length zero and the central barrier, B, completely separates the two halves of the table. In this case $\tilde{P}$ consists of two disjoint tori. The points which map to B in these two tori correspond to two circles which we denote by S_L and S_R corresponding to the left and right chambers. The first return maps on these circles are rotations which we denote by τ_θ. Reflection through the center barrier defines a symmetry of this system which interchanges the two chambers and interchanges S_L and S_R. Let ρ denote this involution. The maps ρ and τ_θ commute. Now let us consider the case when the size of the gap is positive. This has the effect of "coupling" the billiard flows in the two chambers. The points in the circles which correspond to the gap give us two intervals $I_L \subset S_L$ and $I_R \subset S_R$. These intervals are interchanged by the reflection. The first return map is described as follows. If p is not in I_R or I_L then p returns to $\tau_\theta(p)$. If p is in I_R or I_L then p returns to $\rho(\tau_\theta(p))$ and hence it jumps to the other circle. The resulting dynamical system can be viewed as a skew product built over a rotation of the circle with a two point fiber. A precise criterion for unique ergodicity of such maps was given by Veech in [V1]. In particular this criterion can be used to show that there are minimal non-ergodic directional flows for most values of the gap length.

§4. Billiard Techniques: Minimality and Closed Orbits

It might appear from the previous section that the study of polygonal billiards is another form of the study of interval exchange transformations. If we focus only on the flow in a fixed direction then this is largely true. The distinction between the areas appears when we ask not what behavior can occur for various directional flows (our question 1) but rather what behavior is typical (our question 2). Three of the significant problems are to understand the set of directions in which the flow is minimal, the set of directions in which we have periodic billiard orbits and the set of directions in which the flow is uniquely ergodic.

The first result concerns minimal directions.

Theorem 4.1. [ZK]. *The set of directions for which the directional flow is not minimal is countable.*

As we have seen in §3 this involves showing that there are only countably many vertex connections. This follows from the fact that there can be at

most one vertex connection in each relative homotopy class of paths between vertices. This result is valid for all translation structures.

The second question we consider is whether closed billiard trajectories exist. We might ask whether (as in the result of [ZK]) we can use the association between curves and homotopy classes to construct closed trajectories. We might imagine picking a homotopically nontrivial curve on the surface M and shrinking it to obtain a curve of minimal length. This can be done and the resulting curve will be a geodesic in the sense of metric spaces, that is to say that locally it will minimize distance between points. When M is the torus the resulting curve will be a periodic trajectory. When M has genus greater than one the resulting curve will usually not be a periodic trajectory. Typically the geodesic will consist of a sequence of geodesic segments running between vertices so that no two successive segments travel in the same direction.

Despite the fact that this "variational" argument does not work the following theorem of Masur shows that periodic billiard trajectories do exist.

Theorem 4.2. [M2]. *There is a periodic trajectory on every translation surface.*

We will discuss the proof of this result in a moment but we first give a sophisticated extension of this result which is also due to Masur. The most important invariant of a closed orbit is its length. We can ask about the number of families of closed trajectories of length less than N. In the case of the square this question reduces to the question of counting points in the plane with pairs of relatively prime integers as coordinates. The number of such points is asymptotic to cN^2 where $c = \pi/\zeta(2) = 6/\pi$. There is a corresponding asymptotic expression for periodic billiard trajectories in the square or any integrable polygon.

Theorem 4.3. [M3]. *For any translation surface there are constants $0 < c_1 \leq c_2$ such that the number of closed geodesics of length less than N is bounded below by $c_1 N^2$ and bounded above by $c_2 N^2$.*

We will sketch a proof of Theorem 4.2 which shows the usefulness of certain techniques that play an important role in much of the theory of rational billiards and general translation structures. The argument we sketch here follows the logic of Masur's original proof but it replaces the Teichmüller space techniques with techniques based on the geometry of translation structures. A good reference for the techniques we use is [MS1].

We begin by observing that there is a geometric criterion for the existence of a closed geodesic. Let us call a a subset of M isometric to the product of a circle and an interval a *cylinder*. If we fix the genus of M, the number of vertices and the area then the only way for a translation structure on M to have large diameter is for M to contain a long cylinder (that is to say a cylinder where the interval factor is long). (See [MS1] for the simple proof.) If M contains a cylinder then it contains a family of periodic billiard trajectories.

Let M be a surface with a translation structure of area one with no closed billiard trajectories. If M has diameter larger than some constant D then it contains a periodic trajectory and we are done. The important observation is that we can vary the translation structure on M to produce a family of new metrics and apply the geometric argument above to any of these translation structures, not just the original translation structure. We think of a translation structure as being given by an atlas of charts $\{\phi_j\}$ where $\phi_j : U_j \to \mathbf{R}^2$. If we are given a linear map $\alpha : \mathbf{R}^2 \to \mathbf{R}^2$ we can use it to construct a new atlas of charts $\{\phi_j \circ \alpha\}$. Let us denote by $\alpha(M)$ the new translation structure on M. If we take $\alpha \in SL(2, \mathbf{R})$ then $\alpha(M)$ also has area 1. If $\alpha \in SO(2, \mathbf{R})$ then the new translation structure has the same metric as the original one. In general, however, the metric geometry of the new translation structure will be different from the original metric even though the underlying affine structure is the same. A given curve is a periodic trajectory if it does not contain vertices and if it maps to a straight line in any coordinate chart $\phi_i : U \to \mathbf{R}^2$. This is a property of the affine geometry of M which is independent of the metric geometry of M. So if a curve is a periodic trajectory for $\alpha(M)$ it is also a periodic trajectory for M. Of course the diameter of M does depend on the metric. So our hypothesis on M implies that D bounds the diameters of all the translation structures $\alpha(M)$.

Our next objective is to choose a linear transformation which will allow us to exploit this bound on diameter. Let e_1 be the shortest edge in M. We can change the translation structure (and hence the metric) to make e_1 as short as we wish at the expense of making perpendicular directions longer. Choose the translation structure so that e_1 has length less than some constant C_1 to be determined later. Let $\alpha_1(M)$ denote this translation structure. Choose a second edge e_2 disjoint from e_1. Since the diameter of the surface $\alpha_1(M)$ is bounded by hypothesis, we can assume that the length of e_2 is bounded above by some constant, D. We then change the translation structure to get a new translation structure $\alpha_2(M)$ with respect to which e_2 is shorter than some constant C_2. We can continue until we run out of disjoint edges. Since the number of disjoint edges in M is bounded above by some constant n we can construct disjoint edges $e_1, \ldots, e_n$. Let $\epsilon > 0$ be given. We can choose constants $C_1, \ldots, C_n$ so that each edge $e_1, \ldots, e_n$ has length less than ϵ with respect to the final translation structure, $\alpha_n(M)$. We will work backwards from C_n to C_1. Let $C_n = \epsilon$. The last change in the translation structure decreases the length of some edge by a factor of D/C_n. It can increase the length of any other curve by at most this factor. If we choose C_{n-1} to be sufficiently small we can insure that even after performing the last alteration of the translation structure the edge e_{n-1} still has length less than ϵ. Continuing in this way we determine constants $C_1 \ll C_2 \ll C_3 \ll \ldots C_n = \epsilon$.

A maximal collection of disjoint edges in M partition M into a collection of triangles and the number of triangles (say m) depends only on the topology of M. Choose ϵ so that a triangle for which all edges have length less than ϵ has area less than $1/m$. Since all edges have length less than ϵ with respect to the translation structure $\alpha_n(M)$, the total area of $\alpha_n(M)$ is less than one. This contradicts our original assumption that M (and therefore $\alpha_n(M)$) has area one. We conclude that it is not possible for all metrics affinely equivalent to our initial metric M to have bounded diameter. This contradiction proves the existence of a closed trajectory.

An analysis of this argument gives something that Masur's original argument did not give. It produces an explicit upper bound on the length of shortest trajectory.

§5. Billiard Techniques: Unique Ergodicity

In the preceding proof we saw the utility of changing the translation structure and hence the Riemannian metric on a surface while preserving the affine structure. In this section we will show that this technique is also useful in analyzing the ergodic properties of the directional flows.

We will begin by describing a method of changing the translation structure on a surface that gives an analog of Rauzy induction. Let us fix a surface M with a translation structure and consider the directional flow in the vertical direction. To analyze this flow via the methods of interval exchanges we choose a transverse interval which we can take to be horizontal and of length one. The first return map to this interval is an interval exchange. The method of Rauzy induction involves considering a sequence of subintervals $I \supset I_1 \supset I_2 \ldots$ and the sequence of first return maps to these intervals. At each stage we rescale the interval I_n and its subintervals by multiplying its length by λ_n so that its has the same length as I. Now we can achieve the rescaling directly by changing the translation structure so that we multiply the lengths in the horizontal direction by λ_n. To preserve the area of the surface M we can rescale in the vertical direction by multiplying by $(\lambda_n)^{-1}$. We can think of the vertical rescaling as changing the speed of the flow so that the average return time remains constant. Let M_n denote this new translation surface.

Let us define

$$g_t = \begin{pmatrix} e^t & 0 \\ 0 & e^{-t} \end{pmatrix}$$

Then g_t is a one-parameter subgroup of $SL(2, \mathbf{R})$ and the surfaces M_n are just the surfaces $g_{t_n}(M)$ where $t_n = \log \lambda_n$. (In the literature g_t is often defined with a different normalization.)

As we will see there is a criterion for unique ergodicity of the vertical flow on M in terms of the geometry of the surfaces $g_t(M)$ for $t \geq 0$. We call the parametrized collection of translation structures, $\{g_t(M)\}$, a *ray* of translation structures. A ray of translation structures is *divergent* if for any $\epsilon > 0$ there is a T so that for $t > T$ the translation surface $g_t(M)$ has a vertex connection of length less than ϵ. If a ray is not divergent then it is *recurrent*.

We will give two examples of the behavior of rays. Assume that the vertical flow is not minimal. As we have seen this implies that M has a vertical vertex connection. If the length of this vertex connection in M is C then the length of this vertex connection with respect to the metric corresponding to the translation structure $g_t(M)$ is $e^{-t}C$. In particular this ray of translation structures is divergent. Let us consider a second example. Take M to be the torus. Let L be a linear map with integral entries which induces an automorphism of M. Assume that L is hyperbolic with eigenvalues λ and λ^{-1} satisfying $\lambda > 1 > \lambda^{-1} > 0$. Now assume that the translation structure on M is chosen so that the expanding eigenvector of L is horizontal and the contracting eigenvector of L is vertical. In this case the map L induces an isomorphism of translation structures between M and $g_{t_0}(M)$ where $t_0 = \log \lambda$. It follows that $g_t(M)$ is isometric to $g_{t+t_0}(M)$ for any t. We can summarize by saying that, unlike the first case, the geometric invariants of $g_t(M)$ are periodic functions of t. When M has genus greater than one then this periodic behavior occurs precisely when the horizontal and vertical foliations of M arise from a pseudo-Anosov diffeomorphism.

Theorem 5.1. [M4]. *If the vertical flow is not uniquely ergodic then the ray $g_t(M)$ is divergent.*

If the vertical flow is minimal but not uniquely ergodic then instead of having a single edge get short as $t \to \infty$, as in our first example, there will be a sequence of different edges so that the first edge gets short and then as it starts to lengthen a second edge gets short and so on.

To deal with flows in directions other than the vertical we can rotate the translation structure to make the direction vertical and then apply the above criterion. Let

$$r_\theta = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$$

If the flow in direction $\pi/2 - \theta$ is not uniquely ergodic then by Masur's criterion the ray $g_t r_\theta(M)$ is divergent. This test for non-unique ergodicity plays a key role in the following:

Theorem 5.2. [KMS]. *For each translation surface the set of directions for which the flow fails to be uniquely ergodic has measure zero.*

We will discuss the proof of this result after giving a corollary to the theorem and an improvement of the result. Recall that the billiard flow on $T(P)$ is never ergodic when P is irrational. Using the previous result and the technique of approximating non-rational tables by rational tables leads to:

Corollary 5.3. [KMS]. *There exist (non-rational) polygonal billiard tables for which the billiard flow is ergodic on $T(P)$.*

The earlier theorem was improved by Masur to show:

Theorem 5.4. [M4]. *The Hausdorff dimension of the set of directions for which the flow is not uniquely ergodic is at most* $1/2$.

An exposition of the proof of the Theorem 5.2 is contained in [A]. The published proofs rely on a criterion for unique ergodicity which is not as powerful as Masur's criterion (Theorem 5.1). Using theorem 5.1 allows the proof to be simplified somewhat.

We will not describe the proof of Theorem 5.2 here but we will explain how questions about the geometry of translation structures arise in the proof. Let M be a surface with a translation structure. For almost every θ we must show that the ray $g_t r_\theta(M)$ is not divergent as $t \to \infty$. The technique of the proof is to show that for t large and ϵ small the set of θ values for which $g_t r_\theta(M)$ has a segment of length less than ϵ has small measure. Fix t large and assume that for some θ the surface $g_t r_\theta(M)$ has a short edge. Call it e. Now as the translation structure changes we can keep track of the length of e with respect to the translation structure $g_t r_\theta(M)$ as a function of θ. The interval of θ values for which the length of e is greater than ϵ is much larger than the interval for which it has length less than ϵ. The problem is that making the shortest curve longer does not rule out the possibility that some other curve may have gotten shorter in the process. This problem does not arise when M is the torus. In the case of the torus when one curve is short all curves that cross it are long. When M has higher genus though it is possible to have many short curves simultaneously. The solution is to focus on a certain class of curves which behave like curves in the torus. These curves have the property that they are short but not crossed by other short curves.

§6. Dynamics on Moduli Spaces

We have seen how questions about billiard dynamics in rational polygons lead to the study of certain translation structures. In this section we will consider the collection of all translation structures and not restrict ourselves to those arising from polygons. If we identify "geometrically equivalent" translation structures on a given surface then the set of these equivalence classes of structures forms a "moduli space". These moduli spaces possess some interesting and useful geometric structures. The operation of changing the translation structure by elements of $SL(2, \mathbf{R})$ gives a group action of $SL(2, \mathbf{R})$ on each moduli space. We have seen in the previous section how the dynamical properties of the billiard flow on M translate into geometric properties of the various translation structures $\alpha(M)$ for $\alpha \in SL(2, \mathbf{R})$. We will see that the investigation of dynamical properties of the group action on moduli space leads to results about the dynamics of billiard flows for general translation structures as well as to some seemingly unrelated results.

Moduli spaces of translation structures also arise in the theory of quadratic differentials and Teichmüller space. The moduli spaces we will describe are called *strata* because they appear in a stratification of the unit tangent bundle to Teichmüller space. The flow g_t defined in §5 is called the *Teichmüller flow* in this context. For those familiar with Teichmüller space this connection is a source of insight and inspiration. Those not familiar with Teichmüller space methods should not be discouraged. The moduli spaces that we consider can be constructed with no reference to complex analysis (cf. [MS1], [V4] and [V6]) and furthermore none of the theorems discussed in this survey *require* complex methods for their proofs. Whether or not one chooses to use complex methods there is no reason not to use the terminology which arises from the study of Teichmüller space.

Let us say that two translation surfaces M and M' are *topologically equivalent* if the surfaces M and M' are homeomorphic and the number and cone angles of the vertices correspond. (In this survey we are not considering half-integral structures. If we were to do so then there would be an additional piece of data.) Let $f : M \to M'$ be a homeomorphism of translation structures that takes vertices to vertices. If f is a smooth map we can think of the derivative as a map from $\mathbf{R}^2$ to $\mathbf{R}^2$. We say that f is *affine* if the derivative is constant. We say that an affine diffeomorphism f is an *equivalence of translation structures* if the derivative is the identity map. When such an f exists we say that the translation surfaces M and M' are *geometrically equivalent*. For a fixed translation surface M let us denote by $\mathcal{M}(M)$ the moduli space of translation structures topologically equivalent to M with area one. We will call this moduli space a *stratum*. The question of precisely which strata are non-empty is answered in [MS2].

The construction of moduli spaces, that is to say the definition of a topology and other structure for the sets defined above, is rather involved. We will limit ourselves to a description of the construction in the simplest case, that of the surface of genus one. Let T be a torus with a translation structure. Let $\tilde{T}$ denote the universal cover of T. The translation structure gives a canonical way to identify $\tilde{T}$ with $\mathbf{R}^2$. The covering group acts by translation so we can identify it with a lattice $\Lambda \subset \mathbf{R}^2$. The moduli space we want to construct can be identified with the space of lattices in $\mathbf{R}^2$. To build this space we introduce some additional structure. Let us call a translation structure on T together with a choice of a basis of $\pi_1(T)$ a *marked translation structure*. A marked translation structure gives rise to a lattice in $\mathbf{R}^2$ together with a choice of a basis, v and w. Viewing the pair of vectors as a matrix $[vw]$ we can identify the set of marked translation structures with $GL(2, \mathbf{R})$. Now to construct the space of translation structures we analyze the effect of changing the marking on the space of marked translation structures. The mapping class

group of the torus can be identified with the group $GL(2, \mathbf{Z})$. This group acts transitively on the sets of markings. Thus the space of translation structures can be identified with $GL(2, \mathbf{R})/GL(2, \mathbf{Z})$. If we restrict ourselves to translation structures that give the torus area 1 then appropriate moduli space is $SL(2, \mathbf{R})/SL(2, \mathbf{Z})$. The natural action of $SL(2, \mathbf{R})$ is the action by left multiplication. The moduli space for the torus has a smooth finite measure which is invariant under the action of $SL(2, \mathbf{R})$.

The construction of other moduli spaces is somewhat more complicated but it follows the outlines of this construction (see [V4] or [MS1]). As in the case of the torus there is an action of $SL(2, \mathbf{R})$ on the moduli space and a finite smooth measure μ defined on the moduli space invariant under this action. The orbits of the $SL(2, \mathbf{R})$ action are affinely equivalent translation surfaces. Unlike the case of the torus the action of $SL(2, \mathbf{R})$ is not transitive. A second distinction between the higher genus moduli spaces and that of the torus is that the higher genus moduli spaces are not connected in general, but have a finite number of components. This phenomenon was first discovered by Rauzy in the context of interval exchanges.

The moduli space for the torus can be identified with the unit tangent bundle of the modular surface. The Teichmüller flow g_t is just the geodesic flow on the modular surface (though with our normalization of g_t, geodesics travel at twice the usual speed). This geodesic flow is one of the classic examples in dynamical systems. (See Chapter 7 section 5 of this volume for a discussion.) A number of authors have described interesting connections between the geodesic flow on the modular surface and continued fraction expansions. One very elegant way of making this connection is described by Arnoux in [A3] where he uses the fact that each point in $SL(2, \mathbf{R})/SL(2, \mathbf{Z})$ can be interpreted as a flow on the torus. It is a classic result that the geodesic flow for the modular surface is ergodic. The corresponding ergodicity results for the higher genus moduli spaces were proved in general by Veech (certain important cases were established in [M1] and [Re]).

Theorem 6.1. [V4]. *The flow g_t is ergodic on each component of each stratum.*

The flow g_t is uniformly hyperbolic in the genus one case. In the higher genus case it is non-uniformly hyperbolic with respect to the natural measure μ (see [V4] and [KZ]). (For a discussion of non-uniform hyperbolicity see Ch. 7 section 2.8 of this volume) This is a key ingredient in a result of Veech which counts closed orbits for the flow g_t. As we have seen in §5, closed orbits of the Teichmüller flow on $\mathcal{M}(M)$ correspond to pseudo-Anosov diffeomorphisms with prescribed numbers and types of singular points. Let $N(t)$ denote the number of primitive conjugacy classes of pseudo-Anosov diffeomorphisms with the same singularities as M.

Theorem 6.2. [V4].

$$\liminf_{t\to\infty} \frac{\log N(t)}{t} \geq 6g - 6 + 2n$$

where g is the genus of M and n is the number of singular points.

A corollary to this theorem is the existence of pseudo-Anosov diffeomorphisms of all possible topological types (that is to say all possible patterns of topological data) [see MS2].

Considering the $SL(2, \mathbf{R})$ action on moduli space also leads to information about "generic" translation structures.

Theorem 6.3. [MS1]. *For each component of each higher genus moduli space there is a $\delta > 0$ so that for almost every translation surface M in moduli space the set of directions for which the flow is not ergodic has positive Hausdorff dimension.*

The ergodicity of the flow g_t implies that the Hausdorff dimension of the set of non-ergodic directions is constant almost everywhere on each component of each stratum.

Theorem 6.3 shows that there are translation structures of all possible topological types with large sets of non-ergodic directions. The result does not show that there are polygonal billiard tables with this property because the set of translation structures arising from polygonal billiard tables has measure zero in the space translation structures. On the other hand it would certainly be interesting to know to what extent rational billiard tables behave like "generic" translation structures.

Every interval exchange transformation arises as the first return map for some translation structure. Thus the Theorem 6.3 has implications for interval exchange transformations.

Corollary 6.4. [MS1]. *Consider the simplex of interval exchange transformations with a given irreducible permutation not equivalent to a rotation. The subset of the simplex corresponding to not ergodic interval exchange transformations has codimension strictly less than one.*

§7. The Lattice Examples of Veech

While the results we have described in this survey give a great deal of information about general billiard flows they do not in general allow us to say, for a given polygon and a given direction, what the behavior of the directional flow is. In this section we will focus on polygons where such a precise description of the dynamics is possible.

Let us return to a discussion of the billiard flow for the square. In this case each directional flow has one of two types of behavior and we have a straightforward criterion for deciding which type of behavior occurs. If the slope of the direction is rational then all non-singular orbits are closed. If the slope of the direction is irrational then all non-singular orbits are uniformly distributed. Recall that polygonal tables which have only these two types of behavior are said to have the dichotomy property. In this section we will consider polygons where the "dichotomy property" holds and where there is an explicit description of which directions have which behavior.

The list of such "well behaved" tables starts with the integrable polygons. These are precisely the polygons P for which $\tilde{P}$ is the torus. The list of well behaved polygons was extended by Gutkin in his construction of "almost integrable" billiard tables [Gu1]. This class of polygons includes the regular hexagon and it includes polygons all of whose sides are horizontal or vertical and for which the coordinates of vertices are all rational. This list of well behaved polygons was extended by Veech. Veech showed that the list contains not just the regular polygons with 3, 4 and 6 sides but *every* regular polygon.

To describe Veech's result more precisely let us introduce some terminology. If M is a translation surface let $\Gamma(M)$ be the group of affine automorphisms of M. That is to say that Γ consists of homeomorphisms of M which take singularities to singularities and are differentiable with constant derivative away from the singularities. If we consider only orientation preserving maps then the derivative of such a map is an element of $SL(2, \mathbf{R})$. The derivative gives a homeomorphism D from Γ to $SL(2, \mathbf{R})$. We say that M has the *lattice property* if the image of Γ is a lattice in $SL(2, \mathbf{R})$. We say that a polygon P has the lattice property if the translation surface $\tilde{P}$ has the lattice property.

Any polygon that tiles the plane by reflection has the lattice property because $\tilde{P}$ is the torus, T, and $\Gamma(T) = SL(2, \mathbf{Z})$ which is a lattice in $SL(2, \mathbf{R})$. For the almost integrable polygons P the surface $\tilde{P}$ is not the torus but it is a branched cover of the torus where the branch points have rational coordinates. In this case $D\Gamma$ is commensurable to $SL(2, \mathbf{Z})$. Veech's examples also have the lattice property but they differ from the previous examples in that the group Γ is not commensurable to $SL(2, \mathbf{Z})$. In this sense Veech's examples represent a significant new phenomenon in the study of polygonal tables.

Theorem 7.1. [V5], [V7]. *The regular n-gon has the lattice property. A right triangle with one angle of the form π/n has the lattice property. An isosceles triangle with smallest angle of the form $2\pi/n$ has the lattice property.*

The statement that lattice polygons have easily described dynamics is contained in the following theorem.

Theorem 7.2. [V5]. *Lattice examples have the dichotomy property. Furthermore the directions in which all orbits are closed are just the directions fixed by parabolic elements of $D\Gamma$.*

When $D\Gamma$ is commensurable to $SL(2, \mathbf{Z})$ the directions fixed by parabolics are just the rational directions. When $D\Gamma$ is not commensurable to $SL(2, \mathbf{Z})$ these directions are mostly not rational [B]. In addition to identifying directions of closed trajectories it is also possible to analyze the growth rate of closed trajectories (cf. Theorem 4.3).

Theorem 7.3. [V5]. *For any lattice surface there is a constant c such that the number of closed geodesics of length less than N is asymptotic to* cN^2.

Furthermore it is possible, as in the case of the torus, to compute precisely the constants c that arise (see [V5], [V7] and [GJ2]).

We have seen that the properties of the billiard flow on M are captured in the behavior of the orbit of M in moduli space under the $SL(2, \mathbf{R})$ action. This orbit is parametrized by $SL(2, \mathbf{R})/\Gamma$. When Γ is a lattice this space has a special structure which is very much like the structure of $SL(2, \mathbf{R})/SL(2, \mathbf{Z})$. In particular we can identify it with the tangent space to a complete hyperbolic surface of finite area. Any such surface decomposes into a compact piece and a finite number of "cusps". According to Masur's criterion divergent rays correspond to geodesics which eventually remain in a single cusp. Geodesics that remain in a single cusp have a very special form and these correspond to translation structures all of whose vertical trajectories are periodic. It is in this way that the dichotomy property follows from the structure of the $SL(2, \mathbf{R})$ orbit.

Veech's discovery raises the question of whether it will be possible to describe explicitly the dynamics of the billiard flow for other rational tables. An initial question to ask is: Which polygon's have the lattice property? Ward [W] and Vorobets[Vo2] discovered some additional examples among rational triangles and proved that certain specific triangles do not have the lattice property. Kenyon and Smillie [KS] show that among right triangles the examples of Veech are the only ones that have the lattice property. They also analyze a large number, (10^{11}), of acute rational triangles and show that other than Veech's examples only three of these triangles have the lattice property.

There are many open questions related to the description of the billiard flow in explicit rational polygons. If the dynamical behavior of typical translation structures can be taken as a guide to the dynamical behavior of typical rational billiards we would expect that most tables possess non-ergodic directions. On the other hand it seems that no one has yet constructed a single example of a non-ergodic direction in an acute triangle.

It would be very interesting to know about the structure orbits of the $SL(2, \mathbf{R})$ action on moduli space. Is there some analog of Ratner's solution of Raghunathan's conjecture (cf. article by Dani in this volume) which would characterize orbit closures?

As we mentioned in the introduction, rational billiards have been considered as test cases for questions involving quantum chaos in part because their

dynamics is close to the dynamics of integrable systems. As we have seen the dynamics of the lattice examples are particularly well behaved. This motivates the following question of Sinai: Do rational polygons with the lattice property have distinctive quantum mechanical behavior?

Bibliography

[A1] P. Arnoux, Échanges d'intervalles et flots sur les surfaces. (French) Ergodic theory (Sem., Les Plans-sur-Bex, 1980) (French), pp. 5–38, Monograph. Enseign. Math., 29, Univ. Genève, Geneva, 1981.

[A2] P. Arnoux, Ergodicité générique des billards polygonaux (d'après Kerckhoff, Masur, Smillie). Séminaire Bourbaki, Vol. 1987/88. Astérisque No. 161–162 (1988), Exp. No. 696, 5, 203–221 (1989).

[A3] P. Arnoux, Le codage du flot géodésique sur la surface modulaire. Enseign. Math. (2) 40 (1994), no. 1–2, 29–48.

[BKM] C. Boldrighini M. Keane, F. Marchetti, Billiards in polygons, Ann. of Prob. 6, 1978 532–540.

[B] M. Boshernitzan, Billiards and rational periodic directions in polygons, American Math. Monthly 99, 1992, 522–529.

[EG] C. J. Earle, F. P. Gardiner, Teichmüller disks and Veech's F-structures. *Extremal Riemann surfaces* 165–189 Contemp. Math., 201 AMS Providence, RI, 1997.

[FK] R. Fox and R. Kershner, Concerning the transitive properties of geodesics on a rational polyhedron, Duke Math J 2, 1936, 147–150.

[Ga] G. Galperin Non-periodic and not everywhere dense billiard trajectories in convex polygons and polyhedrons, Comm. Math. Phys. 91, 1983 187–211.

[GKT] G. Galperin, T. Kruger S. Troubetzkoy, Local instability of orbits in polygonal and polyhedral billiards, Comm. Math. Phys.

[Gu1] E. Gutkin, Billiards in almost integrable polyhedral surfaces. *Erg. Th. & Dyn. Syst.* **4** (1984), 569–584.

[Gu2] E. Gutkin Billiards in polygons, Physica D **19**, 1986 311–333.

[Gu3] E. Gutkin, Billiards in polygons: survey of recent results. J. Statist. Phys. 83 (1996), no. 1–2, 7–26.

[GT] E. Gutkin and S. Troubetzkoy Directional flows and strong recurrence for polygonal billiards. International Conference on Dynamical Systems (Montevideo, 1995), 21–45, Pitman Res. Notes Math. Ser., 362, Longman, Harlow, 1996.

[GJ1] E. Gutkin and C. Judge, The geometry and arithmetic of translation surfaces with applications to polygonal billiards, *Math. Res. Lett.* **3** (1996) 391–403.

[GJ2] E. Gutkin and C. Judge, Affine mappings of translation surfaces: geometry and arithmetic, preprint (1997).

[HW] Hardy, G. H.; Wright, E. M. An introduction to the theory of numbers. Fifth edition. The Clarendon Press, Oxford University Press, New York, 1979.

[Ka] Katok, A., In variant measures for the flows on oriented surfaces, Dokl. Akad. Nauk SSSR 211, 775–778 (1973) (Russian) English translation: Sov. Math. Dokl. 14, 1104–1108.

[K] M. Keane, Non-ergodic interval exchange transformations, Israel J. Math. 26 (1977), 188–196.

[KS] R. Kenyon, J. Smillie, Billiards on rational-angled triangles, to appear in Commentarii Math. Helv.

[Ke] Kerckhoff, S. P. Simplicial systems for interval exchange maps and measured foliations. Ergodic Theory Dynamical Systems 5 (1985), no. 2, 257–271.

[KMS] S. Kerckhoff, H. Masur, J. Smillie, Ergodicity of Billiard Flows and Quadratic Differentials, Ann. of Math. **124** 1986 p. 293–311.

[KN] H. Keynes and D. Newton, A minimal non-uniquely ergodic interval exchange transformation, Math. Z. **148** (1976), 101–105.

[KS] D. König and A. Szücs, Mouvement d'un point abandonné à l'intérieur d'un cube, Rendiconti del circulo matematico di Palermo, 36 (1913) 79–90.

[KZ] M. Kontsevich and A. Zorich, Lyapunov exponents and Hodge theory, IHES preprint (1997).

[M1] H. Masur, Interval exchange transformations and measured foliations, Ann. of Math. **115** (1982), 169–200.

[M2] H. Masur, Closed trajectories for quadratic differentials with an application to billiards, Duke Math. J. **53** (1986), 307–314.

[M3] H. Masur, The growth rate of trajectories of a quadratic differential. Ergodic Theory Dynamical Systems 10 (1990), no. 1, 151–176.

[M4] H. Masur, Hausdorff dimension of the set of nonergodic foliations of a quadratic differential. Duke Math. J. 66 (1992), no. 3, 387–442.

[MS1] H. Masur, J. Smillie. Hausdorff dimension of sets of nonergodic measured foliations *Annals of Math.* **134** (1991), 455–543.

[MS2] H. Masur. and J. Smillie. Quadratic differentials with prescribed singularities and pseudo-Anosov diffeomorphisms. Comment. Math. Helv. 68 (1993), no. 2, 289–307.

[Ra] G. Rauzy, Échanges d'intervalles et transformations induites, Acta Arith, 34 (1979), 315–328.

[Re] M. Rees, An alternative approach to the ergodic theory of measured foliations on surfaces, Ergodic Theory and Dynamical Systems, Vol. 1 (1981), 461–488.

[RB] P. J. Richens and M. V. Berry, Pseudointegrable systems in classical and quantum mechanics, *Physica* D **2:3** (1981) 495–512.

[Ta] S. Tabachnikoff, Billiards. Panoramas et Synthèses **1** Soc. Math. France (1995).

[V1] W. Veech, Strict ergodicity in zero dimensional dynamical systems and the Kronecker-Weyl theorem mod 2, Trans. A. M. S. **140** (1969), 1–34.

[V2] W. Veech, Interval exchange transformations, J. Analyse Math., 33 (1978), 222–272.

[V3] W. Veech, Gauss measures for transformations on the space of interval exchange maps. Ann. of Math. (2) 115 (1982), no. 1, 201–242.

[V4] W. Veech, The Teichmüller geodesic flow. Ann. of Math. (2) 124 (1986), no. 3, 441–530.

[V5] W. Veech, Teichmüller curves in moduli space, Eisenstein series and an application to triangular billiards. Invent. Math. 97 (1989), no. 3, 553–583.

[V6] W. Veech, Moduli spaces of quadratic differentials. J. Analyse Math. 55 (1990), 117–171.

[V7] W. Veech, The billiard in a regular polygon. Geom. Funct. Anal. 2 (1992), no. 3, 341–379.

[Vo1] Ya. B. Vorobets, Plane structures and billiards in rational polyhedra. (Russian) *Uspekhi Mat. Nauk* 51 (1996), no. 1(307), 145–146; translation in Russian Math. Surveys 51 (1996), no. 1, 177–178

[Vo2] Ya. B. Vorobets, Plane structures and billiards in rational polygons: the Veech alternative. (Russian) *Uspekhi Mat. Nauk* 51 (1996), no. 5(311), 3–42. translation in Russian Math. Surveys 51:5 (1996), 779–817.

[W] C. Ward, Fuchsian groups and polygonal billiards, Thesis, Rice University 1996.

[ZK] A. Zemlyakov, A. Katok. Topological transitivity of billiards in polygons, *Math. Notes of the USSR Acad. Sci.* **18:2** (1975), 291–300. (English translation in Math. Notes **18:2** (1976) 760–764.)

[Z1] A. Zorich, Asymptotic flag of an orientable measured foliation on a surface. Geometric study of foliations (Tokyo, 1993), 479–498, World Sci. Publishing, River Edge, NJ, 1994.

[Z2] A. Zorich, Finite Gauss measure on the space of interval exchange transformations. Lyapunov exponents. Ann. Inst. Fourier (Grenoble) 46 (1996), no. 2, 325–370.

[Z3] A. Zorich, Deviation for interval exchange transformations, Ergodic Theory Dynam. Systems 17 (1997), no. 6, 1477–1499.