Billiards with polynomial mixing rates

N. Chernov[1] and H.-K. Zhang[1]

June 20, 2005

Abstract

While many dynamical systems of mechanical origin, in particular billiards, are strongly chaotic – enjoy exponential mixing, the rates of mixing in many other models are slow (algebraic, or polynomial). The dynamics in the latter are intermittent between regular and chaotic, which makes them particularly interesting in physical studies. However, mathematical methods for the analysis of systems with slow mixing rates were developed just recently and are still difficult to apply to realistic models. Here we reduce those methods to a practical scheme that allows us to obtain a nearly optimal bound on mixing rates. We demonstrate how the method works by applying it to several classes of chaotic billiards with slow mixing as well as discuss a few examples where the method, in its present form, fails.

AMS classification numbers: 37D50, 37A25

1 Introduction

A billiard is a mechanical system in which a point particle moves in a compact container Q and bounces off its boundary ∂Q. It preserves a uniform measure on its phase space, and the corresponding collision map (generated by the collisions of the particle with ∂Q, see below) preserves a natural (and often unique) absolutely continuous measure on the collision space. The dynamical

[1]Department of Mathematics, University of Alabama at Birmingham;
Email: chernov@math.uab.edu; zhang@math.uab.edu

behavior of a billiard is determined by the shape of the boundary ∂Q, and it may vary greatly from completely regular (integrable) to strongly chaotic.

In this paper we only consider planar billiards, where $Q \subset \mathbb{R}^2$. The dynamics in simple containers (circles, ellipses, rectangles) are completely integrable. The first class of chaotic billiards was introduced by Ya. Sinai in 1970 [21]; he proved that if ∂Q is convex inward and there are no cusps on the boundary, then the dynamics is hyperbolic (has no zero Lyapunov exponents), ergodic, mixing and K-mixing. He called such systems *dispersing billiards*, now they are often called *Sinai billiards*. Gallavotti and Ornstein [14] proved in 1976 that Sinai billiards are Bernoulli systems.

Many other classes of planar chaotic billiards have been found by Bunimovich [3, 4] in the 1970s, and by Wojtkowski [23], Markarian [18], Donnay [13], and again Bunimovich [5] in the 1980s and early 1990s. All of them are proven to be hyperbolic, and some – ergodic, mixing, and Bernoulli.

Multidimensional chaotic billiards are known as well, they include two classical models of mathematical physics – periodic Lorentz gases and hard ball gases, for which strong ergodic properties have been established in a series of fundamental works within the last 20 years, we refer the reader to a recent collection of surveys [15].

However, ergodic and mixing systems (even Bernoulli systems) may have quite different statistical properties depending on the rate mixing (the rate of the decay of correlations), whose precise definition is given below. On the one hand, the strongest chaotic systems – Anosov diffeomorphisms and expanding interval maps – have exponential mixing rates, see a survey [12]. This fact implies the central limit theorem, the convergence to a Brownian motion in a proper space-time limit, and many other useful approximations by stochastic processes that play crucial roles in statistical mechanics.

On the other hand, many hyperbolic, ergodic and Bernoulli systems have slow (polynomial) mixing rates, which cause weak statistical properties. Even the central limit theorem may fail, see again a survey [12]. Such systems are, in a sense, *intermittent*, they exemplify a delicate transition from regular behavior to chaos. For this reason they have attracted considerable interest in mathematical physics community during the past 20 years.

The rates of mixing in chaotic billiards is rather difficult to establish, though, because the dynamics has singularities, which aggravate the analysis and make standard approaches (based on Markov partitions and transfer operators) inapplicable. For planar Sinai billiards, L.-S. Young [24] developed in 1998 a novel method (now known as *Young's tower construction*) to prove exponential (fast) mixing rates. Young applied it to Sinai billiards under two restrictions – no corner points on the boundary and finite horizon. Her method was later extended to all planar Sinai billiards [9] and to more general billiard-like Hamiltonian systems [10, 1].

There are no rigorous results on the decay of correlations for multidimensional chaotic billiards yet, since even the best methods cannot handle a recently discovered phenomenon characteristic for billiards in dimension 3 and higher – the blow-up of the curvature of singularity manifolds [2].

For planar billiards with slow (nonexponential) mixing rates, very little is known. They turned out to be even harder to treat than Sinai billiards. Most published accounts are based on numerical experiments and heuristic analysis, which suggest that Sinai billiards with cusps on the boundary [17] and Bunimovich billiards [22] have polynomial mixing rates. Slow mixing in these models appears to be induced by tiny places in the phase space, where the motion is nearly regular, and where the moving particle can be caught and trapped for arbitrarily long times.

A general approach to the studies of abstract hyperbolic systems with slow mixing rates was developed by Young in 1999 [25], but its application to billiards required substantial extra effort, and it took five more years. Only in 2004 R. Markarian [19] used Young's method to establish polynomial mixing rates for one class of chaotic billiards – Bunimovich stadia. He showed that the correlations for the corresponding collision map decayed as $\mathcal{O}\big(n^{-1}\ln^2 n\big)$.

In this paper we generalize and simplify the method due to Young and Markarian essentially reducing it to one key estimate that needs to be verified for each chaotic billiard table. This estimate, see (6.2) in Section 6, has explicit geometric meaning and can be checked by rather straightforward (but sometimes lengthy) computations, almost in an algorithm-like manner. We also remark that the method is not restricted to billiards, it is designed for general nonuniformly hyperbolic systems with Sinai-Ruelle-Bowen (SRB) measures.

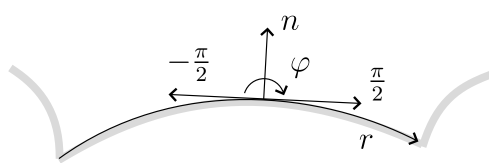
Then we apply the above method to several classes of billiards. These include semi-dispersing billiards in rectangles with internal scatterers, Bunimovich flower-like regions, and "skewed" stadia ("drive-belt" tables). In each case we prove that the correlations decay as $\mathcal{O}\big(n^{-a}\ln^{1+a} n\big)$ for a certain $a > 0$, which depends on the character of "traps" in the dynamics, and is different for different tables. We also show that for some billiard tables our key estimate fails, hence the correlation analysis requires further work.

2 Statement of results

Here we state our estimates on the decay of correlations for several classes of billiard systems. The general method for proving these estimates is described in the next two sections.

First we recall standard definitions [6, 7, 8, 9]. A billiard is a dynamical system where a point moves freely at unit speed in a domain Q (*the table*) and reflects off its boundary ∂Q (*the wall*) by the rule "the angle of incidence equals the angle of reflection". We assume that $Q \subset \mathbb{R}^2$ and ∂Q is a finite union of C^3 curves (arcs). The phase space of this system is a three dimensional manifold $Q \times S^1$. The dynamics preserves a uniform measure on $Q \times S^1$.

Let $\mathcal{M} = \partial Q \times [-\pi/2, \pi/2]$ be the standard cross-section of the billiard dynamics, we call $\mathcal{M}$ the *collision space*. Canonical coordinates on $\mathcal{M}$ are r and φ, where r is the arc length parameter on ∂Q and $\varphi \in [-\pi/2, \pi/2]$ is the angle of reflection, see Fig. 1.



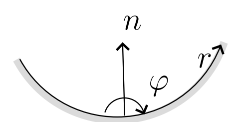


Fig. 1: Orientation of r and φ

The first return map $\mathcal{F} : \mathcal{M} \to \mathcal{M}$ is called the *collision map* or the *billiard map*, it preserves smooth measure $d\mu = \cos\varphi\, dr\, d\varphi$ on $\mathcal{M}$.

Let $f, g \in L^2_\mu(\mathcal{M})$ be two functions. *Correlations* are defined by

$$\mathcal{C}_n(f, g, \mathcal{F}, \mu) = \int_{\mathcal{M}} (f \circ \mathcal{F}^n)\, g\, d\mu - \int_{\mathcal{M}} f\, d\mu \int_{\mathcal{M}} g\, d\mu \tag{2.1}$$

It is well known that $\mathcal{F} : \mathcal{M} \to \mathcal{M}$ is *mixing* if and only if

$$\lim_{n\to\infty} \mathcal{C}_n(f, g, \mathcal{F}, \mu) = 0 \qquad \forall f, g \in L^2_\mu(\mathcal{M}) \tag{2.2}$$

The rate of mixing of $\mathcal{F}$ is characterized by the speed of convergence in (2.2) for smooth enough functions f and g. We will always assume that f and g

are Hölder continuous or piecewise Hölder continuous with singularities that coincide with those of the map $\mathcal{F}^k$ for some k. For example, the free path between successive reflections is one such function.

We say that correlations decay *exponentially* if

$$|\mathcal{C}_n(f,g,\mathcal{F},\mu)| < \text{const} \cdot e^{-cn}$$

for some $c>0$ and *polynomially* if

$$|\mathcal{C}_n(f,g,\mathcal{F},\mu)| < \text{const} \cdot n^{-a}$$

for some $a>0$. Here the constant factor depends on f and g.

Next we state our results.

Let $\mathcal{R} \subset \mathbb{R}^2$ be a rectangle and $\mathcal{B}_1,\dots,\mathcal{B}_r \subset \operatorname{int}\mathcal{R}$ open strictly convex subdomains with smooth or piecewise smooth boundaries whose curvature is bounded away from zero and such that $\overline{\mathcal{B}}_i \cap \overline{\mathcal{B}}_j = \emptyset$ for $i \neq j$, see Fig. 2 (a). A billiard in $Q = \mathcal{R} \setminus \cup_i \mathcal{B}_i$ is said to be *semi-dispersing* since its boundary is partially dispersing (convex) and partially neutral (flat); the flat part is $\partial\mathcal{R}$.

Theorem 1. *For the above semi-dispersing billiard tables, the correlations (2.1) for the billiard map $\mathcal{F}: \mathcal{M} \to \mathcal{M}$ and piecewise Hölder continuous functions f, g on $\mathcal{M}$ decay as $|\mathcal{C}_n(f,g,\mathcal{F},\mu)| \leq \text{const} \cdot (\ln n)^2/n$.*

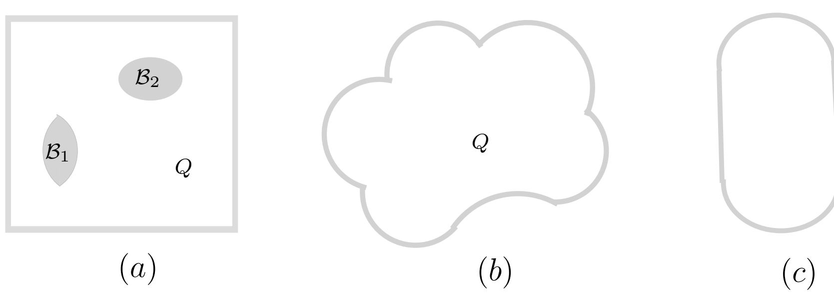


Fig. 2: Slow mixing billiard tables

Next, let $Q \subset \mathbb{R}^2$ be a domain with a piecewise smooth boundary

$$\partial Q = \Gamma = \Gamma_1 \cup \cdots \cup \Gamma_r$$

such that each smooth component $\Gamma_i \subset \partial Q$ is either convex inward (dispersing) or convex outward (focusing). Assume that the curvature of every dispersing component is bounded away from zero and infinity. Also assume that every focusing component Γ_i is an arc of a circle such that there are no points of ∂Q on that circle or inside it, other than the arc Γ_i itself, see Fig. 2 (b). Billiards of this type were introduced by Bunimovich [4] who established their hyperbolicity and ergodicity.

We make two additional assumptions: (i) if two dispersing components Γ_i, Γ_j have a common endpoint, then they are transversal to each other at that point (no cusps!); (ii) every focusing arc $\Gamma_i \subset \partial Q$ is not longer than half a circle. We also assume that the billiard table is generic to avoid certain technical complications, see Section 8.

Theorem 2. *For the above Bunimovich-type billiard tables, the correlations (2.1) for the billiard map $\mathcal{F} : \mathcal{M} \to \mathcal{M}$ and piecewise Hölder continuous functions f, g on $\mathcal{M}$ decay as $|\mathcal{C}_n(f, g, \mathcal{F}, \mu)| \le \text{const} \cdot (\ln n)^3/n^2$.*

In fact, the boundary ∂Q may have flat components as long as there is an upper bound on the number of consecutive reflections off the flat components. For example, this is the case when ∂Q has a single flat component or exactly two nonparallel flat components. Then the above theorem remains valid. However, if the billiard particle can experience arbitrarily many consecutive collisions with flat boundaries, then, generally speaking, even the ergodicity cannot be guaranteed, let alone mixing or decay of correlations.

Lastly, we consider a special case – a stadium. It is a convex domain Q bounded by two circular arcs and two straight lines tangent to the arcs at their common endpoints, see Fig. 2 (c). We distinguish between a "straight" stadium, whose flat sides are parallel, see Fig. 2 (c), and a "skewed" stadium, whose flat sides are not parallel, we call them *drive-belt tables*, see Fig. 10 (a). Notice that drive-belt tables contain an arc longer than a half circle, so they do not satisfy the assumption (ii) of the previous theorem (or the assumptions made by Markarian in [19]). Again billiards of this type were introduced by Bunimovich [4] who established their hyperbolicity and ergodicity.

Theorem 3. *For both types of stadia, the correlations (2.1) for the billiard map $\mathcal{F} : \mathcal{M} \to \mathcal{M}$ and piecewise Hölder continuous functions f, g on $\mathcal{M}$ decay as $|\mathcal{C}_n(f, g, \mathcal{F}, \mu)| \le \text{const} \cdot (\ln n)^2/n$.*

The same bound on correlations for straight stadia is already obtained by Markarian [19], but we include it here for the sake of completeness.

3 Correlation analysis

Here we present a general method for estimating correlations in hyperbolic dynamical systems. It is based on Young's recent results [24, 25] and their extensions by Markarian [19] and one of us [9].

Let $\mathcal{F} : \mathcal{M} \to \mathcal{M}$ be a hyperbolic map acting on a Riemannian manifold $\mathcal{M}$ with or without boundary. We assume that $\mathcal{F}$ preserves an ergodic Sinai-Ruelle-Bowen (SRB) measure μ, see [24, 9] for definitions and basic facts.

Young [24, 25] found sufficient conditions under which correlations for the map $\mathcal{F}$ decay exponentially. We only sketch Young's method here, skipping many technical details, in order to focus on the main condition, see (3.1) below.

The key element of Young's construction is a "horseshoe" Δ_0 (a set with a hyperbolic structure, often called a rectangle; it is obtained by intersection of a family of unstable manifolds with a family of stable manifolds). By iterating points $x \in \Delta_0$ under the map $\mathcal{F}$ until they make proper returns to Δ_0, see definitions in [24], Young constructs a tower, Δ, in which the rectangle Δ_0 constitutes the first level (the base). The induced map $\mathcal{F}_\Delta$ on the tower Δ moves every point one level up until it hits the ceiling upon which it falls onto the base Δ_0. In the end, the tower Δ is identified (mod 0) with the manifold $\mathcal{M}$, and the map $\mathcal{F}_\Delta$ with $\mathcal{F}$.

Now, for $x \in \Delta$, let $R(x; \mathcal{F}, \Delta_0) = \min\{k \geq 1 : \ \mathcal{F}_\Delta^k(x) \in \Delta_0\}$ denote the "return time" of the point x to the base Δ_0. The tower has infinitely many levels, hence R is unbounded. Young proves that if the probability of long returns is exponentially small, then correlations decay exponentially fast. Precisely, if

$$\mu\big(x \in \Delta : \ R(x; \mathcal{F}, \Delta_0) > n\big) \leq \text{const} \cdot \theta^n \qquad \forall n \geq 1 \tag{3.1}$$

where $\theta < 1$ is a constant, then $|\mathcal{C}_n(f, g, \mathcal{F}, \mu)| < \text{const} \cdot e^{-cn}$ for some $c > 0$.

Therefore, if one wants to derive an exponential bound on correlations for a particular map, one needs to construct a "horseshoe" Δ_0 and verify the tail bound (3.1). The latter involves all the iterates of the map $\mathcal{F}$. To simplify this verification one of us (NC) proposed [9] sufficient conditions, which involve just one iterate of the map $\mathcal{F}$ (or one properly selected power $\mathcal{F}^m$, see the next section), that imply the tail bound (3.1), even without constructing a horseshoe Δ_0. We present those conditions fully in the next section.

Young later extended the results of [24] to cover system with slow mixing rates. She proved [25] that if

$$\mu\big(x \in \Delta: \ R(x; \mathcal{F}, \Delta_0) > n\big) \leq \text{ const} \cdot n^{-a} \qquad \forall n \geq 1 \tag{3.2}$$

where $a > 0$ is a constant, then correlations decay polynomially:

$$|\mathcal{C}_n(f, g, \mathcal{F}, \mu)| < \text{ const} \cdot n^{-a} \tag{3.3}$$

Again, a direct verification of the polynomial tail bound (3.2) in specific systems involves all the iterations of the map $\mathcal{F}$ as well as a construction of Δ_0, and thus might be difficult. There are no known simplifications, like the one mentioned above, that would reduce this problem to just one iterate of the map $\mathcal{F}$ and avoid constructing Δ_0. However, there is a roundabout way proposed by Markarian [19] which simplifies the analysis, even though it leads to a slightly less than optimal bound on correlations. We describe Markarian's approach next.

Suppose one can localize places on the manifold $\mathcal{M}$ where the dynamics fails to be strongly hyperbolic, and find a subset $M \subset \mathcal{M}$ on which $\mathcal{F}$ has a strong hyperbolic behavior. This means, precisely, that the first return map $F: M \to M$ satisfies all the conditions of Young's paper [24]; in particular, there exists a horseshoe $\Delta_0 \subset M$ for which an analogue of (3.1) holds:

$$\mu\big(x \in M: \ R(x; F, \Delta_0) > n\big) \leq \text{ const} \cdot \theta^n \qquad \forall n \geq 1 \tag{3.4}$$

for some $\theta < 1$. Of course, the verification of (3.4) can be done via the simplified conditions of [9], which only requires one iteration of F and avoids the construction of Δ_0. But we still need to obtain (3.2) for some $a > 0$. We can do this by considering the return times to M under the original map $\mathcal{F}$, i.e.

$$R(x; \mathcal{F}, M) = \min\{r \geq 1: \ \mathcal{F}^r(x) \in M\} \tag{3.5}$$

for $x \in \mathcal{M}$. Suppose they satisfy a polynomial tail bound

$$\mu(x \in \mathcal{M}: \ R(x; \mathcal{F}, M) > n) \leq \text{ const} \cdot n^{-a} \qquad \forall n \geq 1 \tag{3.6}$$

where $a > 0$ is a constant.

Remark. The bound (3.6) is equivalent to

$$\mu(x \in M: \ R(x; \mathcal{F}, M) > n) \leq \text{ const} \cdot n^{-a-1} \qquad \forall n \geq 1 \tag{3.7}$$

Indeed, denote $\mu_k = \mu(x \in M\colon R(x;\mathcal{F},M)=k)$. Then

$$\mu(x \in M:\ R(x;\mathcal{F},M) > n) = \sum_{k=n+1}^{\infty} \mu_k$$

and, due to the invariance of μ

$$\mu(x \in \mathcal{M}:\ R(x;\mathcal{F},M) > n) = \sum_{k=n+2}^{\infty} (k-1-n)\,\mu_k$$

Now the equivalence of (3.6) and (3.7) is a simple corollary.

Tail bounds for the return times $R(x;\mathcal{F},M)$ to M are much easier to obtain, in many systems, than those for the return times $R(x;\mathcal{F},\Delta_0)$ to Δ_0 (and there is no need to construct Δ_0). The following theorem was essentially proved in [19], but we provide a proof here for completeness:

Theorem 4. *Let $\mathcal{F} : \mathcal{M} \to \mathcal{M}$ be a hyperbolic map. Suppose $M \subset \mathcal{M}$ is a subset such that the first return map $F : M \to M$ satisfies the tail bound (3.4) for the return times $R(x;F,\Delta_0)$ to a rectangle $\Delta_0 \subset M$. If the return times $R(x;\mathcal{F},M)$ satisfy the polynomial bound (3.6) or, equivalently, (3.7), then*

$$|\mathcal{C}_n(f,g,\mathcal{F},\mu)| < \ \mathrm{const}\cdot(\ln n)^{a+1} n^{-a} \tag{3.8}$$

Proof. For every $n \geq 1$ and $x \in \mathcal{M}$ denote

$$r(x;n,M) = \#\{1 \leq i \leq n:\ \mathcal{F}^i(x) \in M\}$$

and

$$\begin{aligned} A_n &= \{x \in \mathcal{M}\colon\ R(x;\mathcal{F},\Delta_0) > n\},\\ B_{n,b} &= \{x \in \mathcal{M}\colon\ r(x;n,M) > b\ln n\}, \end{aligned}$$

where $b > 0$ is a constant to be chosen shortly. By (3.4),

$$\mu(A_n \cap B_{n,b}) \leq \ \mathrm{const}\cdot n\,\theta^{b\ln n}.$$

Choosing b large enough makes this bound less than const$\cdot n^{-a}$.

To bound $\mu(A_n \setminus B_{n,b})$ we note that points $x \in A_n \setminus B_{n,b}$ return to M at most $b\ln n$ times during the first n iterates of $\mathcal{F}$. In other words, there are $\leq b \ln n$ time intervals between successive returns to M, and hence the longest such interval, we call it I, has length $\geq n/(b \ln n)$. Applying the bound (3.7) to the interval I gives

$$\mu(A_n \setminus B_{n,b}) \leq \text{ const} \cdot n\,(\ln n)^{a+1}/n^{a+1}$$

(the extra factor of n must be included because the interval I may appear anywhere within the longer interval $[1, n]$, and the measure μ is invariant). In terms of Young's tower Δ, we obtain

$$\mu(x \in \Delta :\; R(x; \mathcal{F}, \Delta_0) > n) \leq \text{ const} \cdot (\ln n)^{a+1}\, n^{-a} \qquad \forall n \geq 1 \tag{3.9}$$

This tail bound differs from Young's (3.2) by the extra factor $(\ln n)^{a+1}$. However, as it was explained by Markarian [19], the same argument that Young used to derive (3.3) from (3.2) now gives us (3.8) based on (3.9). This completes the proof of the theorem. □

4 Conditions for exponential mixing

Here we list sufficient conditions on a 2D hyperbolic map $F\colon M \to M$ with a mixing SRB measure μ under which its correlations decay exponentially. These are a 2D version of more general conditions stated in [9]. We also provide comments that will help us apply these conditions to chaotic billiards.

We assume that M is an open domain in a two-dimensional C^∞ compact Riemannian manifold $\mathcal{M}$ with or without boundary. For any smooth curve $W \subset \mathcal{M}$ we denote by ν_W the Lebesgue measure on W (induced by the Euclidean metric). For brevity, $|W| = \nu_W(W)$ will denote the length of W.

4.1 Smoothness. The map F is a C^2 diffeomorphism of $M \setminus \mathcal{S}$ onto $F(M \setminus \mathcal{S})$, where $\mathcal{S}$ is a closed set of zero Lebesgue measure. Usually, $\mathcal{S}$ is the set of points at which F either is not defined or is singular (discontinuous or not differentiable).

Remark. The collision map $\mathcal{F}\colon \mathcal{M} \to \mathcal{M}$ for a billiard table with a piecewise smooth C^r boundary is piecewise C^{r-1} smooth [11]. For billiards considered in this paper, the singularities of $\mathcal{F}$ make a closed set $\mathcal{S}$ that is a finite or countable union of smooth compact curves in the collision space $\mathcal{M}$. On those curves, one-sided derivatives of $\mathcal{F}$ are often infinite. When we construct a first return map $F\colon M \to M$ on a subdomain $M \subset \mathcal{M}$, that subdomain will be bounded by some singularity curves, and it will be always clear that F has similar properties.

We denote by $\mathcal{S}_m = \mathcal{S} \cup \mathcal{F}^{-1}(\mathcal{S}) \cup \cdots \cup \mathcal{F}^{-m+1}(\mathcal{S})$ the singularity set for the map $\mathcal{F}^m$. Similarly, $\mathcal{S}_{-m}$ denotes the singularity set for the map $\mathcal{F}^{-m}$.

4.2 Hyperbolicity. There exist two families of cones C^u_x (unstable) and C^s_x (stable) in the tangent spaces $\mathcal{T}_x M$, $x \in \bar{M}$, such that $DF(C^u_x) \subset C^u_{Fx}$ and $DF(C^s_x) \supset C^s_{Fx}$ whenever DF exists, and

$$|DF(v)| \geq \Lambda |v| \quad \forall v \in C^u_x \quad \text{and} \quad |DF^{-1}(v)| \geq \Lambda |v| \quad \forall v \in C^s_x$$

with some constant $\Lambda > 1$. Here $|\cdot|$ is the Euclidean norm. These families of cones are continuous on $\bar{M}$ and the angle between C^u_x and C^s_x is bounded away from zero. Tangent vectors to the singularity curves $\mathcal{S}_m$ for $m > 0$ must lie in stable cones, and tangent vectors to the singularity curves $\mathcal{S}_{-m}$ must lie in unstable cones.

For any F-invariant probability measure μ', almost every point $x \in M$ has one positive and one negative Lyapunov exponent. Also, almost every point x has one-dimensional local unstable and stable manifolds, which we denote by $W^u(x)$ and $W^s(x)$, respectively.

Remark. The existence of Lyapunov exponents usually follows from the Oseledec theorem, and the existence of stable and unstable manifolds – from the Katok-Strelcyn theorem, see [16]. Both theorems require certain mild technical conditions, which have been verified for virtually all planar billiards in [16]. The hyperbolicity has been proven for all the classes of billiards studied in this paper. It is also known that the tangent vectors to the curves in $\mathcal{S}$ lie in stable cones, and those of the curves in $\mathcal{S}_{-1}$ – in unstable cones.

4.3 SRB measure. The map F preserves an ergodic measure μ whose conditional distributions on unstable manifolds are absolutely continuous. Such a measure is called a Sinai-Ruelle-Bowen (SRB) measure. We also assume that μ is mixing.

Remark. For the collision map of a chaotic billiard system, the natural smooth invariant measure μ is an SRB measure [24]. Its ergodicity and mixing have been proven for all classes of billiards studied in this paper, see references in Section 2.

4.4 Distortion bounds. Let $\Lambda(x)$ denote the factor of expansion on the unstable manifold $W^u(x)$ at the point x. If x, y belong to one unstable manifold W^u such that F^n is defined and smooth on W^u, then

$$\log \prod_{i=0}^{n-1} \frac{\Lambda(F^i x)}{\Lambda(F^i y)} \leq \psi\big(\mathrm{dist}(F^n x, F^n y)\big) \tag{4.1}$$

where $\psi(\cdot)$ is some function, independent of W, such that $\psi(s) \to 0$ as $s \to 0$.

Remark. Since the derivatives of the collision map $\mathcal{F}$ turn infinite on some singularity curves $\mathcal{S}$, the expansion of unstable manifolds terminating on those curves is highly nonuniform. To enforce the required distortion bound, one follows a standard procedure introduced in [7], see also [24, 9]. Let $\mathcal{M}_+$ denote the part of the collision space corresponding to dispersing walls $\Gamma_i \subset \partial Q$, i.e.

$$\mathcal{M}_+ = \{(r, \varphi)\colon\ r \in \Gamma_i,\ \Gamma_i \text{ is dispersing}\}.$$

One divides $\mathcal{M}_+$ into countably many sections (called *homogeneity strips*) defined by

$$H_k = \{(r, \varphi) \in \mathcal{M}_+ \colon \pi/2 - k^{-2} < \varphi < \pi/2 - (k+1)^{-2}\}$$

and

$$H_{-k} = \{(r, \varphi) \in \mathcal{M}_+ \colon\ -\pi/2 + (k+1)^{-2} < \varphi < -\pi/2 + k^{-2}\}$$

for all $k \geq k_0$ and

$$H_0 = \{(r, \varphi) \in \mathcal{M}_+ \colon\ -\pi/2 + k_0^{-2} < \varphi < \pi/2 - k_0^{-2}\}, \tag{4.2}$$

here $k_0 \geq 1$ is a fixed (and usually large) constant. Then, a stable (unstable) manifold W is said to be *homogeneous* if its image $\mathcal{F}^n(W)$ lies either in one homogeneity strip of $\mathcal{M}_+$ or in $\mathcal{M} \setminus \mathcal{M}_+$ for every $n \geq 0$ (resp., $n \leq 0$). It is shown in [7] that a.e. point $x \in \mathcal{M}$ has homogeneous stable and unstable manifolds passing through it. The distortion bounds for homogeneous unstable manifolds were proved in [7, 9]. It is also easy to check that the distortion bounds for the collision map $\mathcal{F}$ imply similar bounds, with the same function $\psi(s)$, for the return map $F\colon M \to M$ on any subdomain $M \subset \mathcal{M}$.

We note that the parts of the collision space $\mathcal{M}$ corresponding to the neutral (flat) and focusing walls do not have to be subdivided into 'homogeneity strips'. First, there are no distortions on the neutral components. Second, the parts corresponding to the focusing components are subdivided by the natural singularity manifolds into sufficiently narrow strips on which distortions are already bounded, as it was shown in [7].

From now on, we will only deal with homogeneous manifolds without even mentioning this explicitly. This can be guaranteed by redefining the collision space, as it is done in [9]: we remove from $\mathcal{M}_+$ the boundaries ∂H_k, thus making $\mathcal{M}_+$ a countable disjoint union of the open homogeneity strips H_k's. (There is no need to subdivide or otherwise redefine $\mathcal{M} \setminus \mathcal{M}_+$, see [9]) Accordingly, the images (preimages) of ∂H_k need to be added to the set $\mathcal{S}_{-1}$ (resp. $\mathcal{S}$). We call them *new singularities*, and refer to the original sets $\mathcal{S}$ and $\mathcal{S}_{-1}$ as *old singularities*. Now the stable and unstable manifolds for the map $\mathcal{F}$ on the (thus redefined) collision space $\mathcal{M}$ will be always homogeneous and the distortion bounds will always hold.

4.5 Bounded curvature. The curvature of unstable manifolds is uniformly bounded by a constant $B \geq 0$.

4.6 Absolute continuity. If W_1, W_2 are two small unstable manifolds close to each other, then the holonomy map $h: W_1 \to W_2$ (defined by sliding along stable manifolds) is absolutely continuous with respect to the Lebesgue measures ν_{W_1} and ν_{W_2}, and its Jacobian is bounded, i.e.

$$1/C' \leq \frac{\nu_{W_2}(h(W_1'))}{\nu_{W_1}(W_1')} \leq C' \tag{4.3}$$

with some $C' > 0$, where $W_1' \subset W_1$ is the set of points where h is defined.

Remark. The properties 4.5 and 4.6 have been verified for planar chaotic billiards considered in this paper [7, 9] and even for more general billiard-like Hamiltonian systems [10].

Before we state the last (and most important) condition, we need to introduce some notation. Let $\delta_0 > 0$. We call an unstable manifold W a δ_0-LUM (local unstable manifold) if $|W| \leq \delta_0$. Let $V \subset W$ be an open subset, i.e. a finite or countable union of open subintervals of W. For $x \in V$ denote by $V(x)$ the subinterval of V containing the point x. Let $n \geq 0$. We call an open subset $V \subset W$ a (δ_0, n)-subset if the map F^n is defined and smooth on V and $|F^n V(x)| \leq \delta_0$ for every $x \in V$. Note that $F^n V$ is then a union of δ_0-LUM's. Define a function $r_{V,n}$ on V by

$$r_{V,n}(x) = \operatorname{dist}(F^n x, \partial F^n V(x)), \tag{4.4}$$

which is the distance from $F^n(x)$ to the nearest endpoint of the curve $F^n V(x)$. In particular, $r_{W,0}(x) = \operatorname{dist}(x, \partial W)$.

4.7 One-step growth of unstable manifolds. There is a small $\delta_0 > 0$ and constants $\alpha_0 \in (0,1)$ and $\beta_0, D, \kappa, \sigma > 0$ with the following property. For any sufficiently small $\delta > 0$ and any δ_0-LUM W denote by $U_\delta^1 = U_\delta^1(W) \subset W$ the δ-neighborhood of the set $W \cap \mathcal{S}$ within W. Then there is an open $(\delta_0, 1)$-subset $V_\delta^1 = V_\delta^1(W) \subset W \setminus U_\delta^1$ such that $\nu_W\big(W \setminus (U_\delta^1 \cup V_\delta^1)\big) = 0$ and $\forall \varepsilon > 0$

$$\nu_W\big(r_{V_\delta^1,1} < \varepsilon\big) \le 2\alpha_0\varepsilon + \varepsilon\beta_0\delta_0^{-1}|W| \tag{4.5}$$

$$\nu_W\big(r_{U_\delta^1,0} < \varepsilon\big) \le D\delta^{-\kappa}\varepsilon \tag{4.6}$$

and

$$\nu_W\big(U_\delta^1\big) \le D\delta^{\sigma} \tag{4.7}$$

We comment on these conditions after stating the theorem proved in [9]:

Theorem 5 ([9]). *Let F satisfy the assumptions 4.1–4.7. Then (a) there is a horseshoe $\Delta_0 \subset M$ such that the return times $R(x; F, \Delta_0)$ satisfy the tail bound (3.4). Thus, (b) the map $F : M \to M$ enjoys exponential decay of correlations.*

Remark. It is shown in [9], Proposition 10.1, that if the conclusions (a) and (b) of this theorem hold for some power F^m, $m \ge 2$, of the map F, then they hold for F itself. Therefore, it will be enough to verify the condition 4.7 for the map F^m with some $m \ge 2$ (of course, $\mathcal{S}$ should then be replaced by $\mathcal{S}_m$, $r_{V,1}$ by $r_{V,m}$, etc.). This option will be important when the condition (4.5) fails for the map F but holds for some power F^m, see below. In this case we do not have to worry about (4.6) and (4.7), since they will hold for all F^m, $m \ge 1$, see the next section.

5 Simplified conditions for exponential mixing

Here we simplify the conditions (4.5)–(4.7) and reduce them to one inequality that needs to be checked for every particular class of chaotic billiards.

First, the condition (4.6) always holds for 2D maps, where $\dim W = 1$, as it was shown in [9], and the reason is simple: each connected component of U_δ^1 has length $\geq 2\delta$, thus there are no more than $\delta_0/(2\delta)$ of them, hence $\nu_W\big(r_{U_\delta^1,0} < \varepsilon\big) \leq \delta_0\delta^{-1}\varepsilon$. Thus, (4.6) will hold for all our applications.

The condition (4.7) holds under very mild assumptions on the structure of the set $\mathcal{S}$, which will be always satisfied in our applications:

Assumption (Structure of the singularity set). Assume that for any unstable curve $W \subset M$ (i.e., a curve whose tangent vectors lie in unstable cones) the set $W \cap \mathcal{S}$ is finite or countable and has at most K accumulation points on W, where $K \geq 1$ is a constant. Let x_∞ be one of them and $\{x_n\}$ denote the monotonic sequence of points $W \cap \mathcal{S}$ converging to x_∞. We assume that

$$\operatorname{dist}(x_n, x_\infty) \leq C/n^d \tag{5.1}$$

for some constants $C, d > 0$, i.e. the convergence of x_n to x_∞ is faster than some power function of n.

Lemma 6. *In the notation of 4.7, our assumption (5.1) implies*

$$\nu_W\big(U_\delta^1\big) < 4K\, C^{\frac{1}{1+d}}\, \delta^{\frac{d}{1+d}}, \tag{5.2}$$

which in turn implies (4.7).

Proof. It is obviously enough to consider the case $K = 1$. Let $n_* = C^{\frac{1}{1+d}}\delta^{-\frac{1}{1+d}}$ and note that $U_\delta^1(W)$ is contained in the union of the interval (x_∞, x_{n_*}) and n_* intervals of length 2δ centered at x_i, $1 \leq i \leq n_*$, plus one interval of length 2δ centered at x_∞; thus

$$\nu_W\big(U_\delta^1(W)\big) \leq C/n_*^d + 2(n_* + 1)\delta$$

which easily implies (5.2). □

Lemma 7. *Under our assumption (5.1), the bound (4.7) holds for any power F^m (with D and σ depending on m, of course).*

Proof. We use induction on m. Let W be an unstable manifold and $U_\delta^m(W) \subset W$ denote the δ-neighborhood of the set $W \cap \mathcal{S}_m$ within W. Assume that (4.7) holds, i.e.

$$\nu_W\big(U_\delta^m(W)\big) \leq D\delta^\sigma \tag{5.3}$$

for some constants $D, \sigma > 0$ and any $\delta > 0$. Set

$$\delta_* = \delta^{\frac{d}{(1+d)(1+\sigma)}},$$

then $\nu_W\bigl(U^m_{\delta_*}(W)\bigr) \le D\delta_*^\sigma$ by (5.3). Let $I_1, \dots, I_N$ be the connected components of $W \setminus \mathcal{S}_m$ whose length is $> \delta_*$. Note that $N \le |W|/\delta_* \le \delta_0/\delta_*$. For each $i = 1, \dots, N$ the set $J_i = F^m(I_i)$ is an unstable manifold. By the assumption (5.1), the set $J_i \cap \mathcal{S}$ is at most countable and has $\le K$ accumulation points which satisfy (5.1). Since F^{-m} is smooth and contracting on each J_i, the set $I_i \cap \mathcal{S}_{m+1}$ also is at most countable and has $\le K$ accumulation points which satisfy (5.1). Now by the same argument as in the proof of Lemma 6

$$\nu_W\bigl(U^{m+1}_\delta(I_i)\bigr) \le 4KC^{\frac{1}{1+d}}\,\delta^{\frac{d}{1+d}},$$

for each i. Observe that

$$U^{m+1}_\delta(W) \subset U^m_{\delta_*}(W) \bigcup \biggl(\bigcup_{i=1}^N U^{m+1}_\delta(I_i)\biggr),$$

thus

$$\begin{aligned}\nu_W\bigl(U^{m+1}_\delta(W)\bigr) &\le D\delta_*^\sigma + \text{const} \cdot \delta_*^{-1}\delta^{\frac{d}{1+d}} \\ &\le \text{const} \cdot \delta^{\frac{d\sigma}{(1+d)(1+\sigma)}},\end{aligned}$$

which proves (4.7) for the map F^{m+1}. □

Remark S. We need to extend Lemma 6 to a slightly more general type of singularities. Suppose $\mathcal{S}_1$ and $\mathcal{S}_2$ are two sets such that each satisfies the above assumption (5.1), and $\mathcal{S} = \mathcal{S}_1 \cup F^{-1}(\mathcal{S}_2)$. Then the argument used in the proof of Lemma 7 shows that $\mathcal{S}$ satisfies (4.7). Furthermore, in this case $\mathcal{S}_m$ satisfies (4.7) for every $m \ge 1$, due to the same argument.

We now turn to the most important condition (4.5). Let W_i denote the connected components of $W \setminus \mathcal{S}$ and

$$\Lambda_i = \min_{x \in W_i} \Lambda(x)$$

the minimal local expansion factor of the map F on W_i. We note that due to the distortion bounds

$$\forall x, y \in W_i \qquad e^{-\psi(\delta_0)} \le \frac{\Lambda(x)}{\Lambda(y)} \le e^{\psi(\delta_0)}, \tag{5.4}$$

and so due to the smallness of δ_0 the expansion factor is in fact almost constant on each W_i.

Lemma 8. *The condition (4.5) is equivalent to*

$$\liminf_{\delta_0\to 0} \sup_{W\colon |W|<\delta_0} \sum_i \Lambda_i^{-1} < 1, \tag{5.5}$$

where the supremum is taken over unstable manifolds W, and Λ_i, $i \geq 1$, denote the minimal local expansion factors of the connected components of $W \setminus \mathcal{S}$ under F.

We call (5.5) a *one-step expansion estimate* for F.

Proof. We prove that (5.5) implies (4.5), the converse implication will be self-evident in the end (and the converse will not be used in practical applications anyway).

We choose $\delta_0 > 0$ so that

$$\alpha_1 \colon = \sup_{W\colon |W|<\delta_0} \sum_i \Lambda_i^{-1} < 1.$$

Let $W_i^1 = W_i \setminus \overline{U_\delta^1(W)}$ be the open subinterval of W_i obtained by removing from W_i the δ-neighborhood of its endpoints (of course, if $|W_i| < 2\delta$, then $W_i^1 = \emptyset$), and let $W^1 = \cup_i W_i^1$. It is easy to see that $\forall \varepsilon > 0$

$$\nu_W(r_{W^1,1} < \varepsilon) \leq \sum_i 2\varepsilon\Lambda_i^{-1} = 2\alpha_1\varepsilon \tag{5.6}$$

We note that this estimate is almost exact, since the expansion factor is almost constant on each W_i due to (5.4).

Next, to obtain an open $(\delta_0, 1)$-subset $V_\delta^1 \subset W^1$ as required in (4.5), we need to subdivide the intervals $W_i^1 \subset W^1$ into subintervals whose F-images are shorter than δ_0. More precisely, let us divide each curve $F(W_i^1)$ whose length exceeds δ_0 into $k_i + 1$ equal subintervals, where $k_i = [|F(W_i^1)|/\delta_0]$. If $|F(W_i^1)| \leq \delta_0$, then we set $k_i = 0$ and leave W_i^1 unchanged. Then the union of the preimages of the ε-neighborhoods of the new k_i partition points on the curves $F(W_i^1)$ has ν_W-measure bounded above by

$$3\varepsilon \sum_i k_i\, \Lambda_i^{-1} \leq 3\varepsilon\, \delta_0^{-1} \sum_i |F(W_i^1)|\, \Lambda_i^{-1} \leq 4\varepsilon\, \delta_0^{-1}\, |W|$$

where we increased the numerical coefficient from 3 to 4 in order to incorporate the factor $e^{\psi(\delta_0)}$ resulting from the distortion bounds (5.4). Thus we obtain

$$\nu_W\big(r_{V_\delta^1,1} < \varepsilon\big) \leq 2\alpha_1\varepsilon + 4\varepsilon\delta_0^{-1}|W| \tag{5.7}$$

Lemma is proved. $\square$

The one-step expansion estimate (5.5) is geometrically explicit and easy-to-check, but unfortunately, as stated, it fails too often. Indeed, suppose $S \subset \mathcal{S}$ is a singularity curve that divides an unstable manifold W into two components, W_1 and W_2. For (5.5) to hold, we need $\Lambda_1^{-1} + \Lambda_2^{-1} < 1$, which is a fairly stringent requirement (it means, in particular, that either $\Lambda_1 > 2$ or $\Lambda_2 > 2$). What do we do if the hyperbolicity of the system at hand is not that strong, i.e. the expansion of unstable curves is less than twofold? In that case (5.5) fails, and so does (4.5).

In that case, according to the remark after Theorem 5, it will be enough to prove (4.5) for any power F^m. If we apply Lemma 8 to the map F^m we get the following:

Lemma 9. *The analogue of the condition (4.5) for F^m is equivalent to*

$$\liminf_{\delta_0 \to 0} \sup_{W\colon |W|<\delta_0} \sum_i \Lambda_{i,m}^{-1} < 1, \tag{5.8}$$

where the supremum is taken over unstable manifolds W and $\Lambda_{m,i}$, $i \geq 1$, denote the minimal local expansion factors of the connected components of $W \setminus \mathcal{S}_m$ under F^m.

We call (5.8) the *m-step expansion estimate*.

To summarize our discussion, we state another theorem:

Theorem 10. *Let F be defined on a 2D manifold $\mathcal{M}$ and satisfy the basic requirements 4.1–4.6. Suppose the singularity set $\mathcal{S}$ has the structure described by (5.1) or by Remark S, and let (5.8) hold for some m. Then (a) there is a horseshoe $\Delta_0 \subset M$ such that the return times $R(x; F, \Delta_0)$ satisfy the tail bound (3.4). Thus, (b) the map $F : M \to M$ enjoys exponential decay of correlations.*

In our applications, 4.1–4.6 will always hold and the singularity set $\mathcal{S}$ will obviously have the structure described by (5.1) or by Remark S. The only condition we will need to verify is (5.8) for some $m \geq 1$. But still, it involves a higher iteration of F, which might be practically inconvenient, so we will simplify this condition further in the next section.

6 A practical scheme

The verification of (5.8) can be done according to a general scheme that we outline here.

Suppose $\mathcal{S}_m$ is a finite union of smooth compact curves that are uniformly transversal to unstable manifolds (in our applications, the tangent vectors to those curves lie in stable cones ensuring transversality). For an unstable manifold W, let $K_m(W)$ denote the number of connected components of $W \setminus \mathcal{S}_m$. We call

$$K_m = \lim_{\delta_0 \to 0} \sup_{W\colon |W| < \delta_0} K_m(W)$$

the *complexity of the map* F^m. Note that K_m does not depend on the number of the curves in $\mathcal{S}_m$, which may grow rapidly with m, but rather on the maximal number of those curves that meet at any one point of M.

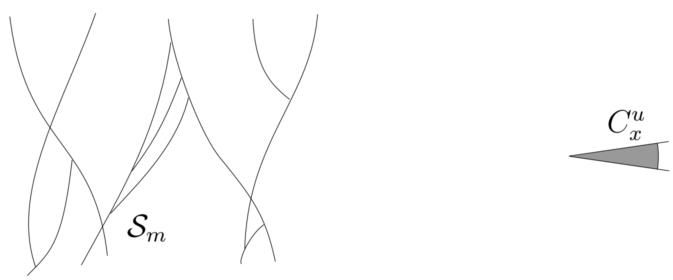


Fig. 3: Singularity curves and unstable cones.

For example, on Fig. 3 a set of 10 curves of $\mathcal{S}_m$ is depicted; to ensure transversality, those curves run in a general vertical direction, while unstable cones (depicted by a shadowed triangle on the right) are assumed to be nearly horizontal; the complexity K_m here equals 3, since sufficiently short unstable manifolds intersect at most 2 curves in $\mathcal{S}_m$.

Lemma 11. *Suppose that for some* $m \geq 1$

$$K_m < \Lambda^m \tag{6.1}$$

where $\Lambda > 1$ *is the minimum expansion factor of unstable vectors. Then (5.8) holds, thus (4.5) holds for* F^m.

Proof. By the chain rule, we have $\Lambda_{m,i} > \Lambda^m$ in (5.8), hence the lemma. $\square$

The condition (6.1) and alike are known as *complexity bounds* in the literature. In some cases, the sequence K_m has been proven to grow with m at most polynomially, $K_m = \mathcal{O}(m^A)$ for some $A > 0$, thus (6.1) holds for all large enough m. It is believed that for typical billiard systems K_m grows slowly (at least more slowly than any exponential function), but only in few cases exact proofs are available.

But we have yet another problem. The above arguments are valid if $\mathcal{S}_m$ is a *finite* union of smooth compact curves. However, in many billiards systems (and in all our applications) $\mathcal{S}_m$ consists of *countably* many smooth compact curves, which accumulate at some places in the space M. For example, the boundaries of homogeneity strips introduced in 4.4, converge to two lines $\varphi = \pm\pi/2$, hence their preimages (included in $\mathcal{S}$) converge to the preimages of the lines $\varphi = \pm\pi/2$. Note that since $\mathcal{S}_m$ is closed, the accumulation points of singularity curves belong to $\mathcal{S}_m$ as well (they may be single points or some curves of $\mathcal{S}_m$).

In this case the singularity curves $\mathcal{S}$ of the map F can be usually divided into two groups: a finite number of *primary* curves and finitely or countably many sequences of *secondary* curves, each sequence converges to a limit point on another singularity curve or to another singularity curve in $\mathcal{S}$. In each example, the partition into primary and secondary singularity curves will be quite natural and explicitly constructed.

Now let W be a short unstable manifold that crosses no primary singularity curves but finitely or countably many secondary singularity curves. Then $W \setminus \mathcal{S}$ consists of finitely or countably many components W_i, and we denote by Λ_i, as before, the minimum local expansion factor of W_i under the map F. In all our applications, the accumulation of secondary singularity curves is accompanied by very strong expansion in their vicinities, so that Λ_i are in fact very large. The following is our crucial assumption:

Assumption (on secondary singularities). We have

$$\theta_0 \colon = \liminf_{\delta_0 \to 0} \sup_{W \colon |W| < \delta_0} \sum_i \Lambda_i^{-1} < 1, \tag{6.2}$$

where the supremum is taken over unstable manifolds W intersecting *no primary singularity curves* and Λ_i, $i \geq 1$, denote the minimal local expansion factors of the connected components of $W \setminus \mathcal{S}$ under the map F.

We denote

$$\theta_1 \colon = \max\{\theta_0, \Lambda^{-1}\} < 1.$$

For the primary singularity curves, we assume an analogue of the complexity bound (6.1) as follows. Let $\mathcal{S}_{\mathrm{P}}$ be the union of the primary singularity curves and $\mathcal{S}_{\mathrm{P},m} = \mathcal{S}_{\mathrm{P}} \cup F^{-1}\mathcal{S}_{\mathrm{P}} \cup \cdots \cup F^{-m+1}\mathcal{S}_{\mathrm{P}}$. For an unstable manifold W, let $K_{\mathrm{P},m}(W)$ denote the number of connected components of $W \setminus \mathcal{S}_{\mathrm{P},m}$. i.e. the number of components into which W is divided by *primary singularities only* during the first m iterations of F. We call

$$K_{\mathrm{P},m} = \lim_{\delta_0 \to 0} \sup_{W \colon |W| < \delta_0} K_{\mathrm{P},m}(W)$$

the *complexity of the primary singularities of* F^m. We now make our last assumption:

Assumption (on primary singularities). For some $m \geq 1$

$$K_{\mathrm{P},m} < \theta_1^{-m} \tag{6.3}$$

Theorem 12. *Assume (6.2). If (6.3) holds for some $m \geq 1$, then (5.8) follows, and thus (4.7) is true for the map F^m.*

Proof. Let $W' \subset W$ be a connected component of the set $W \setminus \mathcal{S}_{\mathrm{P},m}$ and W'_j, $j \geq 1$, denote all the connected components of $W' \setminus \mathcal{S}_m$. Denote by Λ'_j the minimum expansion factor of W'_j under the map F^m.

Lemma 13. *In the above notation*

$$\liminf_{\delta_0 \to 0} \sup_{W \colon |W| < \delta_0} \sup_{W' \subset W} \sum_j [\Lambda'_j]^{-1} < \theta_1^m. \tag{6.4}$$

Proof. During the first m iterations of F, the images of W' never cross any primary singularity curves but may be divided by secondary ones into finitely or countably many pieces.

Now the proof goes by induction on m. For $m = 1$, the statement follows from (6.2) if W' intersects $\mathcal{S}$, or from the mere hyperbolicity 4.2 otherwise. Next we assume (6.4) for some m and apply this same argument to each connected component of the set $F^m(W')$ and then use the chain rule to derive (6.4) for $m + 1$. □

Lemma 13 and the assumption (6.3) imply

$$\liminf_{\delta_0 \to 0} \sup_{W \colon |W| < \delta_0} \sum_i \Lambda_{i,m}^{-1} \leq K_{\mathrm{P},m} \theta_1^m < 1$$

which proves the theorem. $\square$

In the following sections, we apply our methods to three classes of chaotic billiards with slow mixing rates.

7 Proof of Theorem 1

Let $\mathcal{R} \subset \mathbb{R}^2$ be a rectangle and $\mathcal{B}_1, \dots, \mathcal{B}_r \subset \operatorname{int} \mathcal{R}$ open strictly convex subdomains with smooth or piecewise smooth boundaries, as described in Theorem 1.

The billiard system in $Q = \mathcal{R} \setminus \cup_i \mathcal{B}_i$ is semidispersing. The collision space can be naturally divided into two parts: *dispersing*

$$\mathcal{M}_+ = \{(r, \varphi)\colon\ r \in \cup_i \, \partial \mathcal{B}_i\}$$

and *neutral*

$$\mathcal{M}_0 = \{(r, \varphi)\colon\ r \in \partial \mathcal{R}\}$$

We set $M = \mathcal{M}_+$ and consider the return map $F\colon M \to M$. Thus we skip all the collisions with the flat boundary $\partial \mathcal{R}$ when constructing F.

The map F can be reduced to a (proper) collision map corresponding to another billiard table by a standard "unfolding" procedure. Instead of reflecting a billiard trajectory at $\partial \mathcal{R}$ we reflect the rectangle $\mathcal{R}$ with all the scatterers $\mathcal{B}_i$ across the side which our billiard trajectory hits. Then the mirror image of the billiard trajectory will pass straight to the new copy of $\mathcal{R}$ (as if the boundary $\partial \mathcal{R}$ were transparent), see Fig. 4.

The mirror copies of $\mathcal{R}$ obtained by successive reflections across their sides will cover the entire plane $\mathbb{R}^2$. The new (unfolded) billiard trajectory will only hit the scatterers $\mathcal{B}_i$ and their images obtained by the mirror reflections. It is clear that those images make a periodic structure in $\mathbb{R}^2$, whose fundamental domain $\mathcal{R}_1$ consists of four adjacent copies of $\mathcal{R}$, see Fig. 4. In fact, $\mathcal{R}_1$ can be viewed as a 2-dimensional torus, with opposite boundaries identified. Thus we obtain a billiards in a table $Q_1 \subset \mathcal{R}_1$ with some internal convex scatterers and periodic boundary conditions. Our map F reduces to the proper collision map in this new billiard table.

The billiard table Q_1 is dispersing since all the scatterers are strictly convex. It has "infinite horizon" (or "unbounded horizon"), because the free paths between successive collisions may be arbitrarily long. We call $x \in M$ an *infinite horizon point* (or, for brevity, IH-point) if its trajectory forms a

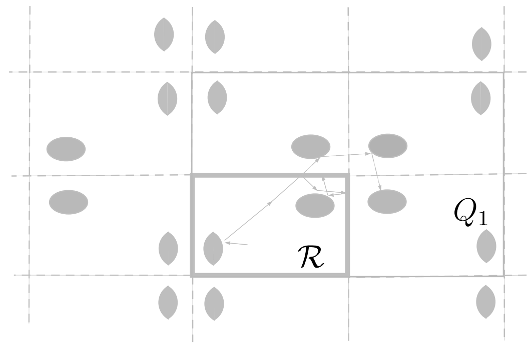


Fig. 4: Unfolding of a billiard trajectory.

closed geodesic on the torus $\mathcal{R}_1$ that only touches (but never crosses) the boundary ∂Q_1, and at all points of contact of this geodesic with ∂Q_1 the corresponding components of ∂Q_1 lie on one (and the same) side of this geodesic. There are at most finitely many IH-points $x \in M$, and we denote them by x_q, $q \geq 1$, they play a crucial role in our analysis.

First, let us assume that the scatterers $\mathcal{B}_i$ have smooth boundaries. Then the billiard in Q_1 is a classical Lorentz gas, also known as Sinai billiard table. The exponential decay of correlations (in fact, both statements (a) and (b) of Theorem 5) for this billiard are proved in [9], Section 8. But the argument in [9] is rather model-specific. Below we outline another argument based on our general scheme described in the previous sections. This outline will also help us treat piece-wise smooth scatterers $\mathcal{B}_i$ later. We mention some standard properties of dispersing billiards and refer the reader to [6, 7, 8, 9] for a detailed account.

The discontinuity curves of the map F (the "old" singularities, in the terminology of Section 4.4) are of two types. There are finitely many long curves dividing M into finitely many large domains. In addition, there are infinite sequences of short singularity curves that converge to the IH-points $x_q = (r_q, \varphi_q) \in M$ (note that $\varphi_q = \pm\pi/2$). The singularity curves divide the neighborhoods of the IH-points into countably many small regions (commonly called *cells*).

The structure of singularity curves and cells in the vicinity of every IH-point x_q is standard – it is shown on the Fig. 5 (a). There is one long curve S_q running from x_q into M and infinitely many almost parallel short curves

$S_{q,n}$, $n \geq 1$, running between S_q and the border ∂M and converging to x_q as n goes to ∞. The dimensions of the components of this structure are indicated on Fig. 5 (a). We denote by $M_{q,n}$ the domain bounded by the curves $S_{q,n}$, $S_{q,n+1}$, S_q and ∂M (this domain is often called n-cell).

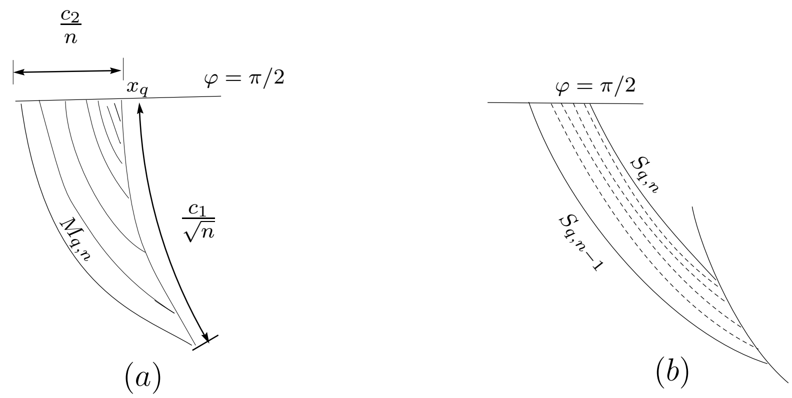


Fig. 5: Singularity curves and cells near an IH-point.

We declare all the short singularity curves $S_{q,n}$ with $n > n_1$ secondary, where n_1 is sufficiently large and will be specified below. All the other old singularity curves are declared primary.

The "new" singularities (the preimages of the boundaries of the homogeneity strips H_k, see Section 4.4) consist of infinite sequences of curves that converge to the old singularities. In particular, every curve $S_{q,n}$ is a limit curve for an infinite sequence of "new" singularity curves, see Fig. 5 (b). All the new singularity curves are declared secondary.

It is easy to check that the distance from the preimage of ∂H_k to the nearest curve $S_{q,n}$ is of order $(nk^2)^{-1}$, which implies the condition (5.1).

It is known that the complexity K_m of the primary singularities of the map F^m grows at most linearly, i.e. $K_m \leq C_1 + C_2 m$, where $C_1, C_2 > 0$ are constants (see [6], Section 8). This means that no more than $C_1 + C_2 m$ singularity curves of $\mathcal{S}_{P,m}$ meet at any one point $x \in M$. This implies (6.3) for any $\theta_1 < 1$ and sufficiently large m.

It remains to verify the main assumption (6.2). Suppose first that a short unstable manifold W crosses new singularity curves converging to a primary old curve. Then W is divided into countably many pieces $W_k = W \cap F^{-1}(H_k)$. The expansion of W_k under F is bounded below by $\Lambda_k \geq Ck^2$, where $C > 0$

is a constant, see [9], Equation (7.3). Thus

$$\sum_{k_0}^{\infty} \Lambda_k^{-1} \le \sum_{k_0}^{\infty} (Ck^2)^{-1} \le 2\,C^{-1}k_0^{-1},$$

which can be made $< 1/2$ by choosing k_0 large enough, say $k_0 > 4\,C^{-1}$.

Next, let a short unstable manifold W intersect some (or all) secondary old curves $S_{q,n}$ with some $q \ge 1$ and $n_1 \le n < \infty$, and near each of them it intersects infinitely many new singularity lines converging to the old one. For each $n \ge n_1$ denote by W_n the piece of W between $S_{q,n}$ and $S_{q,n+1}$, and let $W_{n,k} = W_n \cap F^{-1}(H_k)$ for $n \ge n_1$ and $k \ge k_0$ denote the connected components of $W \setminus \mathcal{S}$. It is important to observe that the image $F(W_n)$ only intersects homogeneity strips H_k with $k \ge \chi n^{1/4}$, where $\chi > 0$ is a constant (see [9], page 544). The expansion of $W_{n,k}$ under F is bounded below by $\Lambda_{n,k} \ge Cnk^2$, where $C > 0$ is a constant, see [9], Equation (8.2). Thus we have

$$\begin{aligned}\sum_{n=n_1}^{\infty} \sum_{k=\chi n^{1/4}}^{\infty} \Lambda_{n,k}^{-1} &\le \sum_{n=n_1}^{\infty} \sum_{k=\chi n^{1/4}}^{\infty} (Cnk^2)^{-1} \\ &\le \sum_{n=n_1}^{\infty} 2\,C^{-1}\chi^{-1}n^{-5/4} \\ &\le 10\,C^{-1}\chi^{-1}n_1^{-1/4},\end{aligned}$$

which can be made $< 1/2$ by choosing n_1 large enough, say $n_1 > [20\,C^{-1}\chi^{-1}]^4$.

This concludes the verification of (6.2). By Theorem 12 and the remark after Theorem 5, the assumptions (4.5)–(4.7) hold, hence the return map $F\colon M \to M$ has exponential mixing rates by Theorem 5.

Lastly, in order to determine the rates of mixing for the original collision map $\mathcal{F}\colon \mathcal{M} \to \mathcal{M}$ of the billiard in $Q = \mathcal{R} \setminus \cup_i \mathcal{B}_i$, we need to estimate the return times defined by (3.7). If the short singularity curves $S_{q,n}$ are labelled in their natural order (so that n increases as they approach the limit IH-point x_q), then

$$\{x \in M\colon R(x;\mathcal{F},M) > n\} \subset \cup_q \cup_{i \ge cn} M_{q,i} \tag{7.1}$$

where $c > 0$ is a constant and $M_{q,i}$ denotes the i-cell. This is a simple geometric fact, we leave the verification to the reader.

It is well known [6, 7] (and can be easily seen on Fig. 5) that $\mu(M_{q,n}) = \mathcal{O}(n^{-3})$, because the "width" of $M_{q,n}$ (its r-dimension) is $\mathcal{O}(n^{-2})$, its "height" (the φ-dimension) is $\mathcal{O}(n^{-1/2})$ and the density of the μ measure in the (r, φ) coordinates is $\cos\varphi = \mathcal{O}(n^{-1/2})$. Hence $\mu\big(R(x; \mathcal{F}, M) > n\big) = \mathcal{O}(n^{-2})$. Theorem 4 now implies the required bound on correlations

$$|\mathcal{C}_n(f, g, \mathcal{F}, \mu)| \leq \text{const} \cdot (\ln n)^2/n.$$

To complete the proof of Theorem 1, we need to consider scatterers $\mathcal{B}_i$ with piece-wise smooth boundaries. Now there are two types of IH-points: those whose trajectories intersect (graze) ∂Q_1 at points where ∂Q_1 is smooth we call them IH1-points, and those whose trajectories intersects ∂Q_1 at corner points, as shown on Fig. 6, we call them IH2-points.

The analysis of IH1-points is identical to the one above. The structure of singularity curves and cells in the neighborhood of an IH2-point $x_q = (r_q, \varphi_q)$ is shown on Fig. 6 (b), we refer the reader to [6], Section 4, for more detail. It is important to note that $|\varphi_q| < \pi/2$ and the first few F-images of a small neighborhood of x_p do not intersect ∂M (they "wander" inside M, as it is explained in [6]), so that there is no "new" singularity curves inside the cells $M_{q,n}$ for large enough n.

Fig. 6: Discontinuity curves and cells near an IH2-point x_q.

Following [7], we assume that the condition (6.3) holds for some $m \geq 1$. We remark that (6.3) holds for generic billiard tables, i.e. for an open dense set of billiard tables in the C^3 metric, which one can show by using standard perturbation techniques, but we do not pursue this goal here.

We proceed to verify the main assumption (6.2). Let $\mathcal{K}_{\min} > 0$ denote the minimum curvature of the boundary of the scatters. Let d denote the width (the smaller dimension) of $\mathcal{R}$ and let

$$b = \min_{i \neq j} \big(\mathrm{dist}(\mathcal{B}_i, \mathcal{B}_j),\ \mathrm{dist}(\mathcal{B}_i, \partial\mathcal{R})\big).$$

Let W be a short unstable manifold and denote by W_n the piece of W between $S_{q,n}$ and $S_{q,n+1}$. Unstable manifolds are represented by increasing curves in the $r\varphi$ coordinates. It is easy to see on Fig. 6 (b) that unstable manifolds cannot cross infinitely many singularity curves. In fact, by a simple geometric calculation, there exists a constant

$$1 \leq \alpha \leq 1 + \frac{k_{\min} + \frac{\cos\varphi_q}{2b}}{k_{\min} + \frac{\cos\varphi_q}{d}}$$

such that if W crosses the curves $S_{q,n}$ for all $n \in [n_1, n_2]$, then $n_2 \leq \alpha n_1$. The expansion of W_n under F (in the Euclidean metric in the $r\varphi$ coordinates) is bounded from below by

$$\Lambda_n \geq 1 + Cn,$$

where $C = \mathcal{K}_{\min} d/\cos\varphi_q$, see [9], Equation (6.8). Here we used the fact that the intercollision time for points $x \in W_n$ is at least $d\,n$. Thus we have

$$\begin{aligned}
\sum_{m=n}^{\alpha n} \frac{1}{\Lambda_m} &\leq \frac{1}{C} \sum_{m=n}^{\alpha n} \frac{1}{m} \\
&\leq \frac{1}{C}\ \ln\Big(1 + \frac{C+\eta}{1+C}\Big) \\
&\leq \frac{C+\eta}{C+C^2} \qquad (7.2)
\end{aligned}$$

where $\eta = d/2b$ and we used the fact that $\ln(1+x) < x$ for $x > 0$.

The bound (7.2) is finite, but we need it to be less than one, which only happens if $\eta < C^2$, that is if $\cos^2\varphi_q \leq 2bd\mathcal{K}_{\min}^2$. While this is indeed the case for some billiard tables, it is not hard to enforce the condition (6.2) for all relevant billiard tables by considering a higher iteration of the map F, as it is explained in [7], page 104.

By Theorem 12 and the remark after Theorem 5, the assumptions (4.5)–(4.7) hold, hence the return map $F\colon M \to M$ has exponential mixing rates.

Lastly, in order to determine the rates of mixing for the original collision map $\mathcal{F}\colon \mathcal{M} \to \mathcal{M}$ of the billiard in $Q = \mathcal{R} \setminus \cup_i \mathcal{B}_i$, we need to estimate the return times defined by (3.7). Similarly to (7.1), we now have

$$\{x \in M\colon R(x; \mathcal{F}, M) > n\} \subset \cup_q \cup_{m \geq c_1 n} M_{q,m}$$

where $c_1 > 0$ is a constant.

It is easy to see on Fig. 6 (b) that $\mu(M_{q,n}) = \mathcal{O}(n^{-3})$ for IH2-points, because the "width" of $M_{q,n}$ (its r-dimension) is $\mathcal{O}(n^{-1})$ and its "height" (the φ-dimension) is $\mathcal{O}(n^{-2})$; hence

$$\mu\big(R(x;\mathcal{F},M) > n\big) = \mathcal{O}(n^{-2}).$$

Theorem 4 now implies the required bound on correlations

$$|\mathcal{C}_n(f,g,\mathcal{F},\mu)| \leq \text{ const} \cdot (\ln n)^2/n. \qquad \square$$

Remark. We emphasize that, apart from the construction of the domain M, our analysis consists of two (very unequal) steps: (i) proving the necessary conditions for the exponential decay of correlations for the return map $F\colon M \to M$, and (ii) the estimation of the tail bound for the return times to M. While the rate of the decay of correlations for the original collision map $\mathcal{F}$ is determined solely at step (ii), that step usually involves much less work than step (i). For all our billiard models step (ii) constitutes an elementary geometric exercise. The novelty of our work lies entirely in step (i), on which we concentrate most.

8 Proof of Theorem 2

Let $Q \subset \mathbb{R}^2$ be a billiard table satisfying the assumptions of Theorem 2. Its boundary can be decomposed as $\partial Q = \partial^+ Q \cup \partial^- Q$ so that each smooth component $\Gamma_i \subset \partial^+ Q$ is dispersing and each component $\Gamma_j \subset \partial^- Q$ is focusing (an arc of a circle that is contained in Q). The collision space can be naturally divided into two parts:

$$\mathcal{M}_\pm = \{(r,\varphi) \,:\, r \in \partial^\pm Q\}.$$

As distinct from dispersing billiards, expansion and contraction in $\mathcal{M}_-$ are not uniform. More precisely, the expansion and contraction are weak (per collision) during long sequences of successive reflections in the same focussing boundary component.

These sequences become a "disturbing factor" similar to reflections in neutral (flat) boundary components studied in the proof of Theorem 1, hence the return map $F\colon M \to M$ must be defined so that those sequences are skipped. We set

$$M = \mathcal{M}_+ \cup \{x \in \mathcal{M}_-\colon\ \pi(x) \in \Gamma_i,\ \pi(\mathcal{F}^{-1}x) \in \Gamma_j,\ j \neq i\}, \tag{8.1}$$

where $\pi(x) = r$ denotes the first coordinate of the point $x = (r, \varphi)$. Observe that M includes all collisions at dispersing boundary components but only the *first* collision at every focusing component. We note that in [7, 19] the *last* collision (rather than the first) in every arc is used in the construction of the return map, but that definition of M leads to unpleasant complications in the analysis, which we will avoid here.

There are two ways in which billiard trajectories can experience arbitrarily many reflections inside one focusing component (arc) $\Gamma_i \subset \partial^- Q$. First, if the collisions are nearly grazing ($\varphi \approx \pm\pi/2$), the points x, $\mathcal{F}(x)$, $\mathcal{F}^2(x), \dots$ are close to each other, so that the sequence $\{\mathcal{F}^n(x)\}$ moves slowly along the arc Γ_i until it comes to an end of the arc and escapes. This is possible for arcs of any size.

Second, if $\varphi \approx 0$, then the billiard trajectory runs near a diameter of Γ_i, hits Γ_i on the opposite side and then comes back, so that the points x, $\mathcal{F}^2(x)$, $\mathcal{F}^4(x), \dots$ are close to each other. Then the two sequences $\{\mathcal{F}^{2n}(x)\}$ and $\{\mathcal{F}^{2n+1}(x)\}$ move slowly along the arc Γ_i until one of them finds an opening in Γ_i and escapes. Similarly, the trajectory can run close to a periodic trajectory inside Γ_i of any period $p \geq 2$. All these, however, require the arc Γ_i to be larger than half-circle, which is specifically excluded by assumption (ii) of Theorem 2. So we do not need to deal with these trajectories now, but we will encounter them later.

In the coordinates (r, φ), the set M is the union of rectangles and cylinders corresponding to the dispersing boundary components and parallelograms corresponding to the focusing boundary components, as the one shown on Fig. 7. In each parallelogram, unstable manifolds are represented by decreasing curves in the (r, φ) coordinates, and singularity manifolds – by increasing curves (which is opposite to dispersing billiards). Between successive collisions at the focusing boundary, each unstable manifold first converges (focuses, or collapses), passes through a conjugate (defocusing) point, and then diverges. We refer the reader to [3, 4, 11] for a detailed account of hyperbolicity in Bunimovich billiards. It is important that the conjugate point always lies in the first half of the segment between the consecutive reflections, which guarantees a monotonic expansion of unstable manifolds from collision to collision in a special p-metric on unstable vectors $du = (dr, d\varphi)$, defined by $|du|_p = \cos\varphi \, dr$, see [6, 7] for definitions and details.

Remark. The p-metric has been used for the estimation of correlations by Sinai et al. [7], Young [24], and Markarian [19]. It is more convenient than the Euclidean metric $|du|^2 = (dr)^2 + (d\varphi)^2$ because the expansion of unstable vectors in monotone in the p-metric (for all the classes of billiards treated in this paper). The p-metric is only a pseudo-metric, but it does not degenerate on stable and unstable tangent vectors, since for all such vectors we have a uniform bound

$$|d\varphi/dr| \le \text{const.} \tag{8.2}$$

Also, the p-metric is equivalent to the Euclidean metric in the following sense: for every homogeneous stable or unstable manifold W, any two points $x, y \in W$, any stable or unstable tangent vector $du \in \mathcal{T}_x M$ and any stable or unstable tangent vector $dv \in \mathcal{T}_y M$ we have

$$0 < c_1 \le \frac{|du|_p/|du|}{|dv|_p/|dv|} \le c_2 < \infty \tag{8.3}$$

where $c_1 < c_2$ are some constants for the given billiard table. This bound follows from (8.2) and the fact that $\cos\varphi$ is almost constant on every homogeneous manifold, due to the distortion bounds. Any metric or pseudo-metric $|\cdot|_p$ that satisfies (8.3) can be used instead of the Euclidean metric, as it follows from the arguments presented in [24, 9].

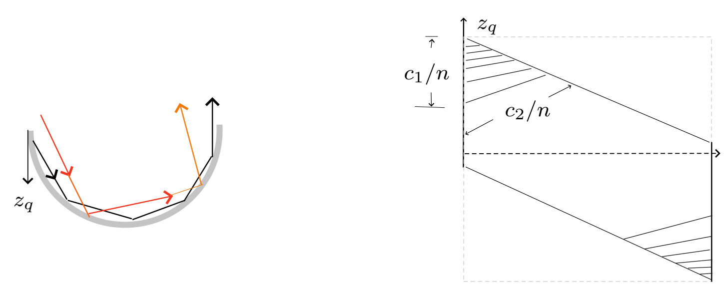


Fig. 7: Sliding trajectories and cell structure of Bunimovich billiard tables

The return map F has infinitely many discontinuity lines, which accumulate in the neighborhood of two vertices (top and bottom) of each parallelogram, see Fig. 7. Each such vertex z_q is a limit point for infinitely many almost parallel straight segment $S_{q,n} \subset \mathcal{S}$ running between two adjacent sides of the parallelogram. They divide the neighborhood of z_q into countably many cells, which we denote, as before, by $M_{q,n}$, $n \ge 1$. Here the cell

$M_{q,n}$ consists of points experiencing exactly n consecutive collisions with the arc (counting the first collision included in M). In fact, for n large enough, those trajectories are more like "sliding" along the arc rather than reflecting off it, hence the corresponding singular points are said to be of *sliding type*, see more details in [6] and [7].

We declare all the short singularity curves $S_{q,n}$ with $n > n_1$ secondary, where n_1 is sufficiently large and will be specified below. All the other singularity curves are declared primary.

Following [7], we assume that the condition (6.3) holds for some $m \geq 1$, see the previous section.

We now proceed to verify the main assumption (6.2). Let $W \subset M_{q,n}$ be an unstable manifold, so that $F = \mathcal{F}^n$ on W. Then the map $\mathcal{F}^{n-1}$ expands W by a factor $\mathcal{O}(n)$, and then $\mathcal{F}$ expands the manifold $\mathcal{F}^{n-1}(W)$ by a factor $\mathcal{O}(n)$, see [7], Section 2.5. Thus, the return map F expands W by a factor $\geq cn^2$, where $c > 0$ is a constant, and we have

$$\sum_{m=n_1}^{\infty} \frac{1}{\Lambda_m} \leq \sum_{m=n_1}^{\infty} \frac{1}{cm^2} \leq \frac{2}{cn_1},$$

which can be made $< 1/2$ by choosing n_1 large enough, say $n_1 > 4/c$.

By Theorem 12 and the remark after Theorem 5, the assumptions (4.5)–(4.7) hold, hence the return map $F\colon M \to M$ has exponential mixing rates.

Lastly, we need to estimate the return times (3.6) to determine the rates of mixing for the original collision map $\mathcal{F}\colon \mathcal{M} \to \mathcal{M}$. The cell $M_{q,n}$ has "width" $\mathcal{O}(n^{-1})$, "height" $\mathcal{O}(n^{-2})$, see Fig. 7, and the density of the measure μ on $M_{q,n}$ is $\mathcal{O}(n^{-1})$, hence $\mu(M_{q,n}) = \mathcal{O}(n^{-4})$. Since

$$\{x \in M\colon\ R(x;\mathcal{F},M) > n\} \subset \cup_q \cup_{m>n} M_{q,m},$$

we get $\mu\big(R(x;\mathcal{F},M) > n\big) = \mathcal{O}(n^{-3})$. Theorem 4 now implies the required bound on correlations

$$|\mathcal{C}_n(f,g,\mathcal{F},\mu)| \leq \text{ const} \cdot (\ln n)^3/n^2.$$

Note that the faster rate of the decay of correlations here (as compared to Theorem 1) is purely due to the smaller cells $M_{q,n}$, and not due to any properties of the return map F. $\square$

9 Proof of Theorem 3

Let $Q \subset \mathbb{R}^2$ be a stadium. Its boundary can be decomposed as

$$\partial Q = \partial^0 Q \cup \partial^- Q,$$

where $\partial Q^0 = \Gamma_1 \cup \Gamma_2$ is the union of two straight sides of Q, and $\partial Q^- = \Gamma_3 \cup \Gamma_4$ is the union of two arcs. The collision space can be naturally divided into focusing and neutral parts:

$$\mathcal{M}_0 = \{(r, \varphi) : r \in \partial^0 Q\}, \qquad \mathcal{M}_- = \{(r, \varphi) : r \in \partial^- Q\}.$$

We define the return map $F\colon M \to M$ on the set

$$M = \{x \in \mathcal{M}_- \colon \pi(x) \in \Gamma_i,\ \pi(\mathcal{F}^{-1}x) \in \Gamma_j,\ j \neq i\}, \tag{9.1}$$

Note that M only contains the *first* collisions with the arcs (the collisions with the straight lines are skipped altogether). In the coordinates (r, φ) the set M is the union of two parallelograms.

First let us consider the straight stadium, see Fig. 2 (c). This model was already handled by Markarian [19], but we present here a simplified proof based on our general method.

The map F can be viewed as a collision map corresponding to another billiard table obtained by "unfolding" the stadium along the straight boundary segments. In other words, instead of reflecting a billiard trajectory at $\partial^0 Q$, we reflect the stadium across its straight boundaries, see Fig. 8.

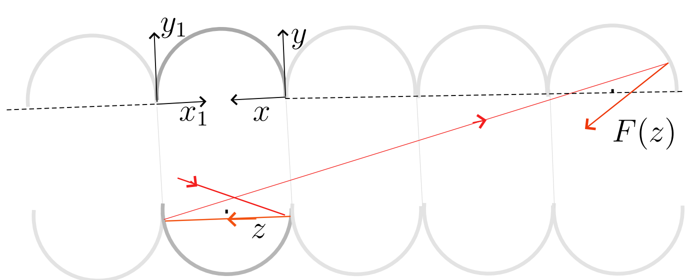


Fig. 8: Unfolding of straight stadium

Then the mirror image of the billiard trajectory will pass straight through ∂Q until it meets the mirror image of the arcs and get reflected there. The mirror copies of the straight stadium obtained by successive reflections about their straight sides make an unbounded strip, see Fig. 8. The new billiard in the unbounded strip has "infinite horizon", since the free path between collisions may be arbitrarily long.

The singularity set $\mathcal{S} \subset M$ of the map F consists of two types of infinite sequences of singularity curves, as shown on Fig. 9. The first type accumulate near the top and bottom vertices of the parallelograms, they are generated by trajectories nearly "sliding" along the circular arcs. These are identical to the ones discuss in the proof of Theorem 2, so we omit them.

The curves of the second type accumulate near the other two vertices of the parallelograms (those lie on the line $\varphi = 0$), they are generated by trajectories experiencing arbitrary many bounces between the two straight sides of the stadium (i.e. by long collision-free flights in the unbounded strip). The structure of those singularity curves is shown on Fig. 9 (b), see [6], Section 6.3, for more details.

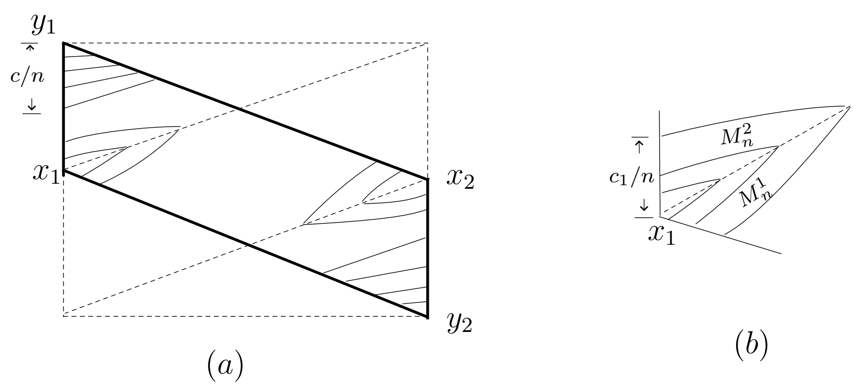


Fig. 9: Discontinuity curves of straight stadium

Actually, there are two types of cells near the vertices x_1 and x_2 in M. The first type contains points mapped (by $\mathcal{F}$) directly to a straight side of the stadium, we denote them by M_n^1, $n \geq 1$; the second type contains points that are mapped by $\mathcal{F}$ onto the same arc first and then to a straight side, we denote them by M_n^2, $n \geq 1$. In Fig. 8, the point z belongs to M_3^2 and $F(z)$ belongs to M_1^1. Points in M_n^1 and M_n^2 experience exactly n reflections off the straight sides before landing on the opposite arc of ∂Q.

To prove the condition (6.3) it is enough to establish the linear growth of the complexity

$$K_{\mathrm{P},m} \le C_1 + C_2 m, \tag{9.2}$$

which is similar to the one obtained for the Lorentz gas in [6], Section 8 (recall our proof of Theorem 1). The proof of (9.2) for the Lorentz gas is based on the continuity of the billiard flow (as it is explained in [10], Lemma 5.2). The flow in the stadium is obviously continuous as well, thus the same argument implies the linear estimate (9.2) for the stadium, we omit details.

We now proceed to verify the main assumption (6.2). Our calculations are based on two known facts mentioned in [7], Section 2.6, which can be verified by direct calculations. First, if an unstable manifold W intersects cells M_n^1 (and M_n^2) with $n_1 < n < n_2$, then

$$n_2 \le 9n_1 + \text{ const.}$$

Second, the expansion factor of F on each piece $W \cap M_n^1$, $n \ge 1$, is $\gtrsim 4n$ and the expansion factor on each piece $W \cap M_n^2$ is $\gtrsim 8n$. Thus we have

$$\sum_{m=n_1}^{n_2} \frac{1}{\Lambda_m} \le \sum_{m=n_1}^{n_2} \Big(\frac{1}{4m} + \frac{1}{8m}\Big) \le \frac{3}{8} \ln 9 < 1.$$

By Theorem 12 and the remark after Theorem 5, the assumptions (4.5)–(4.7) hold, hence the return map $F\colon M \to M$ has exponential mixing rates.

Next we need to estimate the return times (3.6) to determine the rates of mixing for the original collision map $\mathcal{F}\colon \mathcal{M} \to \mathcal{M}$. The n-cells near the top and bottom vertices y_1 and y_2 have measure $\mathcal{O}(n^{-4})$, see the previous section. The n-cells M_n^1 and M_n^2 have measure $\mathcal{O}(n^{-3})$, as one can easily see from Fig. 9 (b) (because the "width" of M_n^1 and M_n^2 (their r-dimension) is $\mathcal{O}(n^{-1})$ and their "height" (the φ-dimension) is $\mathcal{O}(n^{-2})$; note also that $\varphi \approx 0$, hence $\cos\varphi \approx 1$). Thus we get $\mu\big(R(x; \mathcal{F}, M) > n\big) = \mathcal{O}(n^{-2})$. Theorem 4 now implies the required bound on correlations

$$|\mathcal{C}_n(f, g, \mathcal{F}, \mu)| \le \text{ const} \cdot (\ln n)^2/n.$$

Note that there are two types of cells here, and the rate of the decay of correlations is determined by the "worse" (larger) cells M_n^1 and M_n^2 (those near the points x_1 and x_2).

This concludes the proof of Theorem 3 for the straight stadium. $\square$

Next we consider a skewed stadium (or a "drive-belt" table), see Fig. 10 (a). Again, we can unfold the skewed stadium by reflecting it repeatedly along the flat boundaries, as shown on Fig. 10 (b). The new billiard table has similar structure to the Bunimovich type billiard tables of the previous section, but it does not satisfies assumption (ii) of Theorem 2, since it necessarily contains an arc larger than half a circle.

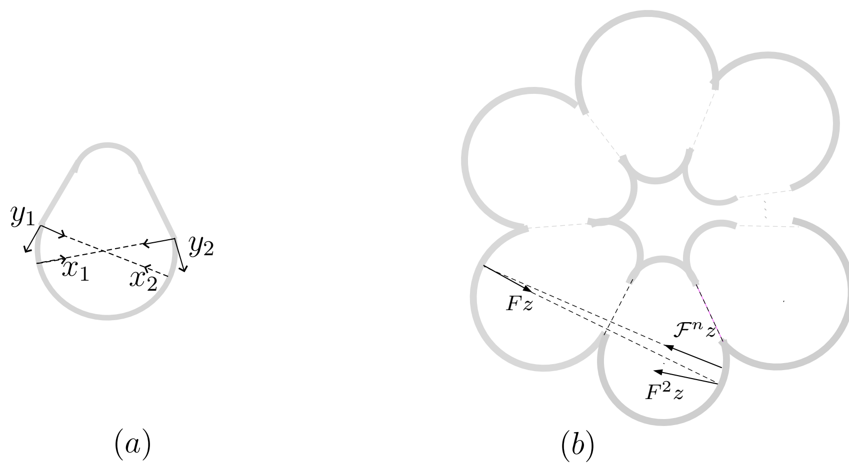


Fig. 10: Unfolding of a skewed stadium

The "unfolding" of skewed stadium also shows that it has "finite horizon", since the free path between successive collisions is uniformly bounded from above, thus there are no cells like M_n^1 and M_n^2 described above, but there is a new, equally influential, type of cells, see below.

In the coordinates (r, φ) the set M is the union of two parallelograms, one corresponds to the smaller arc, the other to the larger arc, we call them *small* and *big* parallelograms, respectively.

The singularity set $\mathcal{S} \subset M$ of the map F consists of two types of infinite sequences of singularity curves. Curves of the first type accumulate near the top and bottom vertices of each parallelogram, they are generated by trajectories nearly "sliding" along the circular arcs. These are identical to the ones discuss in the proof of Theorem 2, so we omit them again.

Curves of the second type accumulate near two points x_1 and x_2 on the slanted sides of the big parallelogram (where those sides intersect the line $\varphi = 0$). There are two parallel long singularity lines S_1 and S_2 starting at x_1 and x_2, respectively, and running into M, and two infinite sequences of almost parallel straight segments $S_{q,n}$, where $q = 1, 2$ and $n \geq 1$, running between S_q and the nearby side of M and converging to x_q as $n \to \infty$, see Fig. 11.

The singularities of this second type are generated by trajectories experiencing arbitrary many collisions with the large arc while running almost along its diameter (this type of trajectories was described in the previous section, but not studied in detail there, because it was ruled out by the assumption (ii) of Theorem 2). Denote by $M_{q,n}$ the n-cell bounded by $S_{q,n}$, $S_{q,n+1}$, S_q and ∂M. The n-cell $M_{q,n}$ consists of points experiencing exactly n collisions with the large arc.

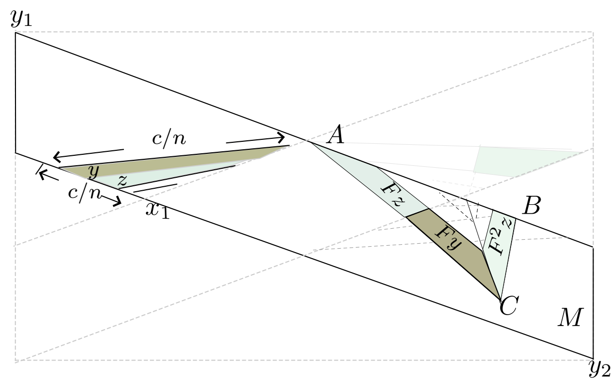


Fig. 11: The cell structure of skewed stadium

There is a peculiar feature of the new cells $M_{q,n}$. Points in the upper half of $M_{q,n}$ near x_q (the dark grey area marked by y on Fig. 11) are mapped by F into the cells $M_{3-q,m}$, $n_1 \leq m \leq n_2$, near the other limit point x_{3-q} (see the dark grey area marked by Fy), and $n_2 \leq 49n_1 + \text{const}$, see below. However, points in the lower half of the cell $M_{q,n}$ (the light grey area marked by z on Fig. 11) are mapped under F into a neighborhood of the other limit point x_{3-q} above or below the line S_{3-q} without crossing any singularity lines there (see the light grey area marked by Fz), and the second iterate of F, as it is illustrated by Fig. 10 (b), maps them into the cells $M_{3-q,m}$, $n_1 \leq m \leq n_2$ with $n_2 \leq 49n_1 + \text{const}$ (see the light grey area marked by F^2z).

The bound $n_2 \leq 49n_1 + \text{const}$ follows from elementary calculations, we only outline them. Suppose the larger arc of ∂Q measures $\pi + \gamma$ radians, $\gamma > 0$, and has radius R. Then the action of the map $\mathcal{F}$ on the part of $\mathcal{M}$ corresponding to this arc satisfies $\mathcal{F}^2(r, \varphi) = (r - 4R\varphi, \varphi)$. Hence, within the cell $M_{q,n}$ we have $\varphi = \frac{\gamma}{8n} + \mathcal{O}(1/n^2)$. Also, the slanted sides of the corresponding parallelogram in M have slope $d\varphi/dr = (2R)^{-1}$. By following the trajectories of points $(r, \varphi) \in M_{q,n}$ one collision after they have left the larger arc we obtain the coordinates of the points A, B, and C shown on Fig. 11: $A(-\frac{R\gamma}{4n}, \frac{\gamma}{8n})$, $B(\frac{R\gamma}{28n}, -\frac{\gamma}{56n})$, $C(\frac{5R\gamma}{4n}, -\frac{7\gamma}{8n})$ (all the r coordinates are given relative to that of x_2, and all these coordinates are exact up to a term $\mathcal{O}(1/n^2)$). Now we see that $n_1 = n/7 + \mathcal{O}(1)$ and $n_2 = 7n + \mathcal{O}(1)$.

Fig. 10 (b) shows the dynamics of the lower half of the cell $M_{q,n}$. Any point z in that part experiences n collisions with the large arc, then crosses a straight side of Q and lands on the adjacent copy of the large arc (this becomes Fz), but then it crosses the same straight side back and lands on the old large arc again, where it starts another long series of m reflections, $n_1 \le m \le n_2$, running nearly along the arc's diameter.

The above analysis implies that if an unstable manifold W intersects cells $M_{q,n}$ with $n_1 \le n_2$, then $n_2 \le 49 n_1 + \text{const}$. Next, by a direct calculation (we omit details) the expansion factor of the map F on the curve $W \cap M_{q,n}$ is $\gtrsim 8n$. However, since $W_n = W \cap M_{q,n}$ consists of two parts (the upper half, call it W_n', and the lower half, call it W_n''), which evolve differently, the number of pieces of W in our estimate is doubled and we get

$$\sum_{m=n_1}^{n_2} \frac{1}{\Lambda_m'} + \frac{1}{\Lambda_m''} \le 2 \sum_{m=n_1}^{n_2} \frac{1}{8m} \eqsim \frac{1}{4} \ln 49 < 1.$$

This finishes the proof of (6.2). Now by Theorem 12 and the remark after Theorem 5, the assumptions (4.5)–(4.7) hold, hence the return map $F\colon M \to M$ has exponential mixing rates.

Lastly, we need to estimate the return times (3.6) to determine the rates of mixing for the original collision map $\mathcal{F} : \mathcal{M} \to \mathcal{M}$. The cells near the top and bottom vertices of the parallelograms of M have measure $\mathcal{O}(n^{-4})$, see the previous section. The n-cells M_n have measure $\mathcal{O}(n^{-3})$, as one can easily see from Fig. 11 (note again that $\varphi \approx 0$, hence $\cos\varphi \approx 1$). Thus we get $\mu\big(R(x; \mathcal{F}, M) > n\big) = \mathcal{O}(n^{-2})$. Theorem 4 now implies the required bound on correlations

$$|\mathcal{C}_n(f, g, \mathcal{F}, \mu)| \le \text{ const} \cdot (\ln n)^2/n.$$

The proof of Theorem 3 is complete. $\square$

10 Other examples and open questions

Here we discuss two types of chaotic billiard tables for which our method fails. This discussion will demonstrate the limitations of the method, in its present form, and indicate directions for future work.

First, let Q be a Bunimovich type table satisfying the assumption (i), but not (ii), of Theorem 2, that is let $\partial^- Q$ contain an arc Γ_i larger than half a circle. Suppose we define the return map $T\colon M \to M$ by (8.1) again, then M will contains a parallelogram M_i corresponding to the arc Γ_i, and the structure of the singularity lines in M_i will be the same as the one shown on Fig. 11. In particular, the expansion factor of F on any unstable manifold $W \cap M_{q,n}$ will be $\sim cn$, where $c > 0$ is a constant.

However, unlike the case of a drive-belt region illustrated on Fig. 11, now some unstable manifolds may intersect infinitely many cells $M_{q,n}$, in particular all n-cells with $n \geq n_1$ for some $n_1 > 0$. In that case

$$\sum_{m=n_1}^{\infty} \frac{1}{\Lambda_m} \sim \sum_{m=n_1}^{\infty} \frac{1}{cm} = \infty, \tag{10.1}$$

and so the condition (6.2) fails. And it fails in a major way, since neither F nor any power F^m can possibly satisfy (6.2).

This failure poses interesting questions. Does this mean that the return map $F\colon M \to M$ has subexponential decay of correlations? If not, can the method be improved to overcome the trouble and establish exponential mixing for F? Or should one seek a different definition of the return map F in order to avoid the trouble? These questions remain open at the moment.

A similar failure occurs for a special modification of the stadium, where Q is bounded by two parallel straight segments and two arcs which are less than half a circle (in that case the arcs have to be transversal to the segments at the intersection points). For the hyperbolicity and ergodicity of this model, see [6], Section 6.3. Suppose we define the return map $F\colon M \to M$ by (9.1), as before. Then one can investigate the singularities of this map and find (see Section 6.3 of [6] for more details) that an unstable manifold W can be broken by $\mathcal{S}$ into infinitely many pieces W_n, $n \geq n_1$, and the expansion factor of the map F on W_n is $\sim cn$, where $c > 0$ is a constant. Hence again we arrive at a divergent series similar to (10.1), so the condition (6.2) fails. It is not clear what this trouble implies for the map F and how to deal with it.

These two example demonstrate the limitations of our algorithm, in its present form. There are other models, not discussed in this paper, to which our scheme probably applies but it requires substantial extra effort. One of them is a dispersing billiard table with cusps (corner points where the adjacent boundary components are tangent to each other). The estimation of mixing rates for this model remains an open problem since the first publication [17] on it in 1983. The other is a dispersing billiard table where the curvature of the boundary vanishes at some points, i.e. where the boundary looks like the graph of function $y = c|x|^\beta$ near $x = 0$ with some $\beta > 2$ (then the curvature vanishes at $x = 0$). Our preliminary calculations show that the rate of mixing is polynomial and its degree depends on β. The work on both models is currently underway.

Another interesting question is how to bound the correlations from below (recall that all our bounds are from above). It is intuitively clear from the results of [25] that the polynomial bounds on correlations we established here (up to the logarithmic factor, though) are sharp, i.e. cannot be improved. However, formal lower bounds on correlations may be hard to obtain. We only point out recent results by O. Sarig in this direction [20].

References

[1] Balint, P. and Toth, I. P., *Correlation decay in certain soft billiards*, Commun. Math. Phys. **243** (2003), 55–91.

[2] Balint, P.; Chernov, N.; Szasz, D. and Toth, I. P., *Geometry of multidimensional dispersing billiards*, Asterisque **286** (2003), 119–150.

[3] Bunimovich L. A., *On billiards close to dispersing*, Math. USSR Sbornik, **23** (1974), 45–67.

[4] Bunimovich, L. A, *On the ergodic properties of nowhere dispersing billiards*, Comm. Math. Phys. **65** (1979), 295–312.

[5] Bunimovich L. A., *On absolutely focusing mirrors*, Lect. Notes Math., **1514**, 62–82, Springer, New York, 1990.

[6] Bunimovich, L. A.; Sinai, Ya. G. & Chernov, N. I. *Markov partitions for two-dimensional hyperbolic billiards,* Russian Math. Surveys **45** (1990) 105–152.

[7] Bunimovich L. A.; Sinai, Ya. G. & Chernov, N. I., *Statistical properties of two-dimensional hyperbolic billiards,* Russian Math. Surveys **46** (1991) 47–106.

[8] Chernov, N., *Entropy, Lyapunov exponents and mean-free path for billiards*, J. Statist. Phys. **88** (1997), 1–29.

[9] Chernov, N., *Decay of correlations in dispersing billiards*, J. Statist. Phys. **94** (1999), 513–556.

[10] Chernov, N., *Sinai billiards under small external forces*, Ann. H. Poincaré **2** (2001), 197–236.

[11] Chernov, N. and Markarian, R., *Introduction to the Ergodic Theory of Chaotic Billiards*, 2nd Ed., IMPA, Rio, Brasil, 2003.

[12] Chernov, N. and Young, L.-S., *Decay of correlations for Lorentz gases and hard balls*, in: Hard Ball Systems and the Lorentz Gas, D. Szasz, ed., Encyclopaedia of Mathematical Sciences **101**, pp. 89–120, Springer, 2000.

[13] Donnay V., *Using integrability to produce chaos: billiards with positive entropy*, Comm. Math. Phys. **141** (1991), 225–257.

[14] Gallavotti, G. and Ornstein, D., *Billiards and Bernoulli schemes*, Comm. Math. Phys. **38** (1974), 83–101.

[15] *Hard Ball Systems and the Lorentz Gas*, D. Szasz, ed., Encyclopaedia of Mathematical Sciences **101**, Springer, 2000.

[16] Katok, A. and Strelcyn, J.-M., *Invariant manifolds, entropy and billiards; smooth maps with singularities*, Lect. Notes Math., **1222**, Springer, New York, 1986.

[17] Machta, J., *Power law decay of correlations in a billiard problem*, J. Statist. Phys. **32** (1983), 555–564.

[18] Markarian, R., *Billiards with Pesin region of measure one*, Comm. Math. Phys. **118** (1988), 87–97.

[19] Markarian, R., *Billiards with polynomial decay of correlations*, Er. Th. Dynam. Syst. **24** (2004), 177–197.

[20] Sarig, O., *Subexponential decay of correlations*, Invent. Math. **150** (2002), 629–653.

[21] Sinai, Ya. G., *Dynamical systems with elastic reflections. Ergodic properties of diepersing billiards,* Russian Math. Surveys **25** (1970) 137–189.

[22] Vivaldi,F.; Casati, G. and Guarneri, I., *Origin of long-time tails in strongly chaotic systems*, Phys. Rev. Lett. **51** (1983), 727–730.

[23] Wojtkowski, M. P., *Principles for the design of billiards with nonvanishing Lyapunov exponents*, Commun. Math. Phys. **105** (1986), 391–414.

[24] Young L.-S., *Statistical properties of systems with some hyperbolicity including certain billiards*, Ann. Math., **147** (1998), 585–650.

[25] Young, L.-S., *Recurrence times and rates of mixing*, Israel J. Math. **110** (1999), 153–188.

Figure captions

1. Orientation of r and φ
2. Slow mixing billiard tables
3. Singularity curves and unstable cones
4. Unfolding of a billiard trajectory
5. Singularity curves and cells near an IH-point
6. Discontinuity curves and cells near an IH2-point x_q
7. Sliding trajectories and cell structure of Bunimovich billiard tables
8. Unfolding of straight stadium
9. Singularity curves and cells near an IH-point
10. Discontinuity curves and cells near an IH2-point x_q
11. The cell structure of skewed stadium