Journal of Statistical Physics, Vol. 122, No. 6, March 2006 (© 2006)
DOI: 10.1007/s10955-006-9036-8

Advanced Statistical Properties of Dispersing Billiards

N. Chernov[1]

Received October 05 2005; accepted January 31 2006
Published Online: March 21, 2006

A new approach to statistical properties of hyperbolic dynamical systems emerged recently; it was introduced by L.-S. Young and modified by D. Dolgopyat. It is based on coupling method borrowed from probability theory. We apply it here to one of the most physically interesting models—Sinai billiards. It allows us to derive a series of new results, as well as make significant improvements in the existing results. First we establish sharp bounds on correlations (including multiple correlations). Then we use our correlation bounds to obtain the central limit theorem (CLT), the almost sure invariance principle (ASIP), the law of iterated logarithms, and integral tests.

KEY WORDS: Sinai billiards, decay of correlations, central limit theorem, invariance principle, law of iterated logarithms.

1. INTRODUCTION

A billiard is a mechanical system in which a point particle moves freely (by inertia) in a compact container $\mathcal{D}$ and bounces off its boundary $\partial\mathcal{D}$. The dynamical properties of a billiard are determined by the shape of $\partial\mathcal{D}$, and they may vary greatly from completely regular (integrable) to strongly chaotic. The first class of chaotic billiards was introduced by Ya. Sinai in 1970, see,[(23)] who considered containers defined by

$$\mathcal{D} = \mathrm{Tor}^2 \setminus \cup_{i=1}^{p} \mathbb{B}_i, \tag{1.1}$$

where Tor^2 denotes the unit 2-torus and $\mathbb{B}_i \subset \mathrm{Tor}^2$ disjoint strictly convex domains (scatterers) with C^3 smooth boundary whose curvature nowhere vanishes. Sinai proved[(23)] that the billiard flows and maps in such domains are hyperbolic, ergodic,

[1] Department of Mathematics, University of Alabama at Birmingham, Birmingham, AL 35294; e-mail: chernov@math.uab.edu.

0022-4715/06/0300-1061/0 © 2006 Springer Science+Business Media, Inc.

and K-mixing. He called these systems *dispersing billiards*, now they are known as *Sinai billiards*.

More advanced ergodic properties where established for Sinai billiards as well: Bernoulliness was proved in(15), Markov partitions were constructed in(4,6), estimates on periodic points were obtained in(6,25). Statistical properties were established fairly recently; these include exponential decay of correlations(9,26), Central Limit Theorem (CLT) and Weak Invariance Principle (WIP), see(5,7). All these facts demonstrate that dispersing billiards are strongly chaotic systems and can be placed in the same category as Anosov and Axiom A diffeomorphisms.

Traditional methods for proving statistical properties are based on Markov partitions and symbolic dynamics(20−22). If a system admits a finite Markov partition (this is the case for Anosov and Axiom A maps), then its symbolic system is a subshift of finite type. Then one usually shows that the corresponding Perron-Frobenius operator (acting on Hölder continuous densities) has a spectral gap, and then derives all the above statistical properties (and more) combining methods of functional analysis and probability theory.[2]

In the case of billiards, however, Markov partitions are never finite, which rendered the symbolic representation inefficient and the Perron-Frobenius operator unhandy. First attempts to investigate statistical properties of dispersing billiards used Markov approximations(5,7) and were technically overcomplicated. They did produce the CLT and WIP but gave only sub-optimal (sub-exponential) estimates on correlations.

Later Young developed(26) a more efficient approach to general hyperbolic maps with singularities, based on *tower construction* (which is a tractable version of countable Markov partitions), and found a way to reemploy the Perron-Frobenius operator. This produced an exponential bound on correlations for dispersing billiards.(9,26) Still, the tower construction is fairly complicated, and the use of the Perron-Frobenius operator is not quite convenient when a parametric family of models is studied, see(12).

Then Young rederived(27) exponential bound on correlations for hyperbolic maps by a different method (borrowed from probability theory) based on coupling of smooth measures, thus bypassing symbolic formalism and the Perron-Frobenius operator altogether. The underlying idea of this new approach is that the images of different smooth measures on the phase space of billiard map are getting closer together, and thus converge to a common limit, the degree of 'closeness' and the speed of convergence are controlled by coupling. Her entire argument was intrinsically dynamical and highly flexible, unlike earlier operator-based proofs.

[2] This is a very efficient approach, but it relies upon highly abstract elements of functional analysis (spectral properties of pseudo-compact operators), which in a sense obscure the dynamical content of the problem.

The elegance of Young's new method was recently demonstrated in[2] where it was adapted to Anosov maps. Dolgopyat further simplified[12][Appendix A] the coupling method by replacing smooth measures on phase space with one-dimensional measures on unstable curves, which made the argument even more transparent and almost elementary. Here we employ Dolgopyat's version of the coupling method.

The paper is organized as follows. Section 2 contains the necessary background on Sinai billiards. Section 3 describes the coupling method and states its key tool—Coupling Lemma (whose proof is given in Appendix). In Section 4 we establish sharp bounds on correlations for (dynamically) Hölder continuous functions. These include bounds on multiple correlations, which follow from our arguments almost automatically, but otherwise are rather hard to establish[11].

In Section 5 we combine our estimates on correlations with a general result by Philipp and Stout[19] to prove various limit theorems: Central limit theorem (CLT), Weak Invariance Principle (WIP), Almost Sure Invariance Principle (ASIP), Law of Iterated Logarithms, and Integral Tests. The last three results are actually new, in the billiard context.

It is interesting to note that all these probabilistic limit theorems follow from sharp bounds on multiple correlations. The earlier proofs of limit theorems (in particular, that of CLT, see[5,7,8,13,26]) used bounds on correlations also, but mainly relied upon some other (stronger) mixing properties of the dynamics. One always wondered if the CLT could be derived solely from correlation bounds. We demonstrate that this is indeed possible: in fact, all our limit theorems formally follow from our bounds on correlations, so that no other mixing properties of the dynamics are necessary.

2. PRELIMINARIES

Here we recall basic facts about dispersing billiards. All of them are well known, see.[6,7,9,10,23,26]

Let $\mathcal{D} \subset \mathrm{Tor}^2$ be a domain defined by (1.1) and $\partial\mathcal{D} = \cup_i \partial\mathbb{B}_i$ its boundary. The billiard particle moves inside $\mathcal{D}$ with constant (unit) speed and bounces off $\partial\mathcal{D}$ according to the classical law "the angle of incidence is equal to the angle of reflection."

The phase space of the billiard system is $\Omega = \mathcal{D} \times S^1$, and the billiards dynamics generates a flow $\Phi^t \colon \Omega \to \Omega$. It is a Hamiltonian (contact) flow, and it preserves Liouville (uniform) measure on Ω.

At every reflection the velocity vector changes by the rule $v^+ = v^- - 2\,\langle v, n\rangle\, n$, where v^+ and v^- refer to the postcollisional and precollisional velocities, respectively, and n denotes the inward unit normal vector to $\partial\mathcal{D}$ at the reflection point $q \in \partial\mathcal{D}$. The family of postcollisional velocity vectors with footpoints on

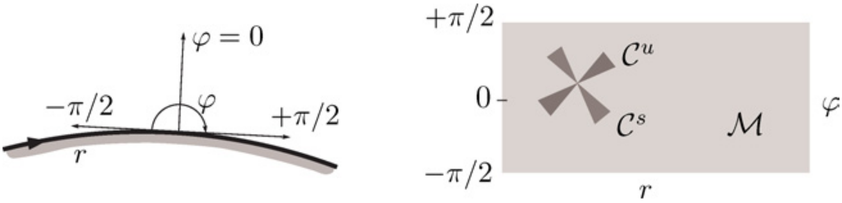

Fig. 1. Orientation of r and φ and the collision space.

$\partial\mathcal{D}$ makes a 2D manifold called the *collision space*

$$\mathcal{M} = \{x = (q, v) \in \Omega \colon\ q \in \partial\mathcal{D}\ \langle v, n\rangle \geq 0\},$$

where $\langle\cdot\rangle$ denotes the scalar product. The billiard flow induces a billiard map (also called collision map) $\mathcal{F}\colon \mathcal{M} \to \mathcal{M}$.

Standard coordinates on $\mathcal{M}$ are the arc length parameter r on the boundary $\partial\mathcal{D}$ and the angle $\varphi \in [-\pi/2, \pi/2]$ between the vectors v and n; the orientation of r and φ is shown on Fig. 1. Note that $\langle v, n\rangle = \cos\varphi$. The space $\mathcal{M}$ is the union of p cylinders on which r is a cyclic ('horizontal') coordinate and φ is a linear ('vertical') coordinate. The map $\mathcal{F}\colon \mathcal{M} \to \mathcal{M}$ preserves smooth measure $d\mu = c_\mu \cos\varphi\ dr\, d\varphi$, where c_μ is the normalizing factor.

For $x \in \mathcal{M}$ denote by $\tau(x)$ the distance of the free path between the collision points we at x and $\mathcal{F}(x)$. The flow Φ^t can be represented as a suspension flow over the map $\mathcal{F}\colon \mathcal{M} \to \mathcal{M}$ under the ceiling function $\tau(x)$. Clearly $\tau(x) \geq \tau_{\min} > 0$, where $\tau_{\min}$ is the minimal distance between the domains $\mathbb{B}_i \subset \mathrm{Tor}^2$. If $\tau(x)$ is bounded above ($\tau(x) \leq \tau_{\max} < \infty$), then the billiard is said to have *finite horizon*. We not not assume finite horizon here.

The billiard map $\mathcal{F}$ is hyperbolic. There is a family of $D\mathcal{F}$-invariant unstable cones $\mathcal{C}^u_x \subset \mathcal{T}_x\mathcal{M}$ and a family of $D\mathcal{F}^{-1}$-invariant stable cones $\mathcal{C}^s_x \subset \mathcal{T}_x\mathcal{M}$. They can be defined so that $c_1 \leq d\varphi/dr \leq c_2$ for all $(dr, d\varphi) \in \mathcal{C}^u_x$ and $-c_2 \leq d\varphi/dr \leq -c_1$ for all $(dr, d\varphi) \in \mathcal{C}^s_x$, where $0 < c_1 < c_2 < \infty$ are constants.

A smooth curve $W \subset \mathcal{M}$ is said to be *unstable* (or *stable*) if at every point $x \in W$ the tangent space $\mathcal{T}_x W$ belongs to the unstable (stable) cone. Note that unstable curves are increasing and stable curves are decreasing in the $r\varphi$ coordinates, and their slopes are bounded away from zero and infinity. Unstable curves are stretched by $\mathcal{F}$, while stable curves are stretched by $\mathcal{F}^{-1}$. Unstable curves correspond to dispersing wave fronts, and stable curves correspond to convergent wave fronts[6,7].

A curve $W \subset \mathcal{M}$ is an *unstable manifold* (or *stable manifold*) if $\mathcal{F}^n(W)$ is an unstable (stable) curve for all $n \leq 0$ (respectively, $n \geq 0$). There exists a (locally unique) unstable (stable) manifold through a.e. point $x \in \mathcal{M}$.

REMARK (ON CONES). As usual in the studies of hyperbolic maps, the construction of invariant cones is rather loose. For example, given a family of unstable cones $\mathcal{C}^u_x$ we can replace it with more narrow cones $\mathcal{C}_x^{u,n} := D\mathcal{F}^n(\mathcal{C}^u_{\mathcal{F}^{-n}x})$. This reduces the class of unstable curves. Ultimately, we can set $n = \infty$, then the cones become lines tangent to unstable manifolds, and unstable curves reduce to unstable manifolds.

For any point $x \in W$ on an unstable (or stable) curve $W \subset \mathcal{M}$ we denote by $\mathcal{J}_W\mathcal{F}^n(x) = \|D_x\mathcal{F}^n(dx)\|/\|dx\|$, $dx \in \mathcal{T}_x W$, the Jacobian of the restriction of the map $\mathcal{F}^n$ to W, at the point x. Then the hyperbolicity means that there are constants $\Lambda = \Lambda(\mathcal{D}) > 1$ and $C = C(\mathcal{D}) > 0$ such that for any unstable curve W^u and any stable curve W^s and all $n \geq 1$ we have $\mathcal{J}_{W^u}\mathcal{F}^n(x) \geq C\Lambda^n$ and $\mathcal{J}_{W^s}\mathcal{F}^{-n}(x) \geq C\Lambda^n$. Note that Λ and C do not depend on W^u or W^s.

In general, the stretching is very nonuniform, as the stretching factor grows to infinity near $\partial\mathcal{M} = \{\cos\varphi = 0\}$. To control distortions of stable and unstable curves under the action of $\mathcal{F}$, it was proposed by Sinai[(7)] to partition them by countably many lines that are parallel to $\partial\mathcal{M}$ and accumulate near $\partial\mathcal{M}$. Let $k_0 \geq 1$ be a large constant. For each $k \geq k_0$ we define two *homogeneity strips* $\mathbb{H}_{\pm k} \subset \mathcal{M}$ by

$$\mathbb{H}_k = \{(r,\varphi): \pi/2 - k^{-2} < \varphi < \pi/2 - (k+1)^{-2}\}$$

and

$$\mathbb{H}_{-k} = \{(r,\varphi): -\pi/2 + (k+1)^{-2} < \varphi < -\pi/2 + k^{-2}\}.$$

We also put $\mathbb{H}_0 = \{(r,\varphi): -\pi/2 + k_0^{-2} < \varphi < \pi/2 - k_0^{-2}\}$. Now $\mathcal{M}$ is divided into homogeneity strips $\mathbb{H}_k$. Denote by $\mathbb{S}_k = \{(r,\varphi): |\varphi| = \pi/2 - k^{-2}\}$ for $|k| \geq k_0$ the boundaries of the homogeneity strips and put $\mathbb{S} = \cup_{|k|\geq k_0}\mathbb{S}_k$. A stable or unstable curve $W \subset \mathcal{M}$ is said to be *weakly homogeneous* if W belongs to one strip $\mathbb{H}_k$.

Now distortions can be characterized as follows. Let W be an unstable curve such that $W_n = \mathcal{F}^{-n}(W)$ is a weakly homogeneous unstable curve for all $0 \leq n \leq N-1$. Then we have the following *distortion bounds*, see[(10)]:

$$C_1 \leq e^{-C|W|^{1/3}} \leq \frac{\mathcal{J}_W\mathcal{F}^{-n}(y)}{\mathcal{J}_W\mathcal{F}^{-n}(z)} \leq e^{C|W|^{1/3}} \leq C_1, \tag{2.1}$$

for every $y, z \in W$ and every $1 \leq n \leq N$; here $C, C_1 > 0$ are constants. Due to time reversibility, similar bounds hold for stable curves.

Next we describe the singularities of $\mathcal{F}$. We denote by $\mathcal{S}_0 = \partial\mathcal{M} = \{\cos\varphi = 0\}$ the boundary of the collision space (it consists of all ‘grazing’ collisions). The map $\mathcal{F}$ lacks smoothness on the set $\mathcal{S}_1 = \mathcal{S}_0 \cup \mathcal{F}^{-1}(\mathcal{S}_0)$ (we call it the *singularity set* for $\mathcal{F}$). In fact, $\mathcal{F}$ is discontinuous on $\mathcal{S}_1 \setminus \mathcal{S}_0$. More generally, the singularity sets for the maps $\mathcal{F}^n$ and $\mathcal{F}^{-n}$ are $\mathcal{S}_n = \cup_{i=0}^n \mathcal{F}^{-i}(\mathcal{S}_0)$ and $\mathcal{S}_{-n} = \cup_{i=0}^n \mathcal{F}^{i}(\mathcal{S}_0)$. For each $n \geq 1$, the set $\mathcal{S}_{-n} \setminus \mathcal{S}_0$ is a finite or countable union of smooth unstable curves (in fact, it is finite for billiards with finite horizon and countable otherwise). Similarly, the set $\mathcal{S}_n \setminus \mathcal{S}_0$ is a finite or countable union of smooth stable curves.

In billiards without horizon, the smooth components of the singularity sets $\mathcal{S}_1$ and $\mathcal{S}_{-1}$ accumulate near finitely many points on $\mathcal{S}_0$ and divide their neighborhoods into 'cells' described in[6,7].

Since, following Sinai, we partition stable and unstable curves by the lines $\mathbb{S}_k$, $|k| \geq k_0$, it is convenient to divide the entire collision space $\mathcal{M}$ into homogeneity strips $\mathbb{H}_k$, $|k| \geq k_0$, and $\mathbb{H}_0$. So we get a new collision space $\mathcal{M}_{\mathbb{H}}$ constructed as a disjoint union of the closures of the $\mathbb{H}_k$'s; note that $\mu(\mathcal{M}_{\mathbb{H}}) = 1$ and $\partial\mathcal{M}_{\mathbb{H}} = \mathbb{S}$.

The map $\mathcal{F}\colon \mathcal{M} \to \mathcal{M}$ naturally acts on the new collision space $\mathcal{M}_{\mathbb{H}}$. The map $\mathcal{F}\colon \mathcal{M}_{\mathbb{H}} \to \mathcal{M}_{\mathbb{H}}$ lacks smoothness on the 'extended' singularity set $\mathcal{S}_1 \cup \mathbb{S} \cup \mathcal{F}^{-1}(\mathbb{S})$. Similarly, the map $\mathcal{F}^n\colon \mathcal{M}_{\mathbb{H}} \to \mathcal{M}_{\mathbb{H}}$ is not smooth on the 'extended' singularity set

$$\mathcal{S}_n^{\mathbb{H}} := \mathcal{S}_n \cup \left(\cup_{m=0}^n \mathcal{F}^{-m}(\mathbb{S})\right).$$

The inverse map $\mathcal{F}^{-n}\colon \mathcal{M}_{\mathbb{H}} \to \mathcal{M}_{\mathbb{H}}$ is not smooth on the 'extended' singularity set

$$\mathcal{S}_{-n}^{\mathbb{H}} := \mathcal{S}_{-n} \cup \left(\cup_{m=0}^n \mathcal{F}^{m}(\mathbb{S})\right).$$

For each $n \geq 1$, the set $\mathcal{S}_{-n}^{\mathbb{H}} \setminus \mathcal{S}_0^{\mathbb{H}}$ is a countable union of smooth unstable curves. Similarly, the set $\mathcal{S}_n^{\mathbb{H}} \setminus \mathcal{S}_0^{\mathbb{H}}$ is a countable union of smooth stable curves. The set $\mathcal{S}_0^{\mathbb{H}}$ consists of parallel lines that are neither stable nor unstable.

For each $n', n'' \geq 0$ the set $\mathcal{M} \setminus (\mathcal{S}_{-n'}^{\mathbb{H}} \cup \mathcal{S}_{n''}^{\mathbb{H}})$ is a countable union of open domains with piecewise smooth boundaries ('curvilinear polygons'), see[6,7]. Moreover, the interior angles made by their boundary components do not exceed π (i.e. those polygons are 'convex', as far as the interior angles are concerned), see[6,7]; some interior angles may be equal to zero.

An unstable curve $W \subset \mathcal{M}$ is said to be a homogeneous unstable manifold (or, briefly, *unstable H-manifold*) if $\mathcal{F}^{-n}(W)$ is a weakly homogeneous unstable curve for every $n \geq 0$. Similarly, a *stable H-manifold* is a curve $W \subset \mathcal{M}$ such that $\mathcal{F}^n(W)$ is a weakly homogeneous stable curve for every $n \geq 0$. For any unstable (stable) H-manifold W we have $|\mathcal{F}^n W| \leq C\Lambda^{-|n|}$ for all $n \leq 0$ (resp., $n \geq 0$); here $|W|$ denotes the length of W.

Let ξ_n^u denote the measurable partition of $\mathcal{M}$ into the connected components of the set $\mathcal{M} \setminus \mathcal{S}_{-n}^{\mathbb{H}}$. Then $\xi^u = \vee_{n\geq 1}\xi_n^u$ is the measurable partition of $\mathcal{M}$ into the (maximal) unstable H-manifolds (see a proof in[10]). Similarly, if ξ_n^s denotes the measurable partition of $\mathcal{M}$ into the connected components of the set $\mathcal{M} \setminus \mathcal{S}_n^{\mathbb{H}}$, then $\xi^s = \vee_{n\geq 1}\xi_n^s$ is the measurable partition of $\mathcal{M}$ into the (maximal) stable H-manifolds. We denote by $W^u(x)$ and $W^s(x)$ the (maximal) stable and unstable H-manifolds passing through the point $x \in \mathcal{M}$.

The conditional measures on unstable H-manifolds $W \subset \mathcal{M}$ are absolutely continuous and their densities ρ_{W^u} satisfy

$$\frac{\rho_W(y)}{\rho_W(z)} = \lim_{n\to\infty} \frac{\mathcal{J}_W \mathcal{F}^{-n}(y)}{\mathcal{J}_W \mathcal{F}^{-n}(z)} \tag{2.2}$$

for $y, z \in W$ (this is a standard formula in the studies of hyperbolic maps that first appeared in(22), see also [18, Theorem 3]). For any unstable H-manifold $W \subset \mathcal{M}$, the unique probability density ρ_W satisfying (2.2) is called the *u-SRB density*, and the corresponding probability measure ν_W on W is called the *u-SRB* measure. The distortion bounds (2.1) imply the following bounds on the u-SRB density, see(3):

$$\left|\frac{d}{dx} \ln \rho_W(x)\right| \le \frac{C}{|W|^{2/3}}, \tag{2.3}$$

where $C = C(\mathcal{D}) > 0$ is a constant. It immediately follows that

$$e^{-C|W(x,y)|^{1/3}} \le \frac{\rho_W(x)}{\rho_W(y)} \le e^{C|W(x,y)|^{1/3}}, \tag{2.4}$$

for all $x, y \in W$; here $W(x, y)$ denotes the segment of the curve W between the points x and y.

For μ-almost every point $x \in \mathcal{M}$ there exist nonvanishing stable and unstable H-manifolds $W^s(x)$ and $W^u(x)$ through x. The point x divides the H-manifold $W^s(x)$ (and $W^u(x)$) into two segments, we denote by $r^s(x)$ (resp., $r^u(x)$) the length of the shorter one. Then have the following estimate (see, e.g.,(10):

$$\mu\{x\colon \min\{r^u(x), r^s(x)\} < \varepsilon\} \le C\varepsilon \tag{2.5}$$

for some constant $C = C(\mathcal{D}) > 0$ and all $\varepsilon > 0$. Moreover, for any stable (or unstable) curve $W \subset \mathcal{M}$ we have

$$m_W\{x \in W\colon r^p(x) < \varepsilon\} \le C\varepsilon \tag{2.6}$$

where $p = u$ if W is stable and $p = s$ if W is unstable, and m_W denotes the Lebesgue measure on W. Note that (2.6) is, in a sense, a local version of (2.5)

The following is known as Sinai's Fundamental Theorem (it is actually a strengthened version of [23, Theorem 6.1]):

Theorem 2.1. *Let $x \in \mathcal{M} \setminus \cup_{n\ge 0} \mathcal{S}_n^{\mathbb{H}}$. Then for any $q > 0$ and $A > 0$ there exists an open neighborhood $\mathcal{U}_x^u \subset \mathcal{M}$ of x such that for any unstable curve $W \subset \mathcal{U}_x^u$*

$$m_W\big(y \in W\colon r^s(y) > A|W|\big) \ge (1-q)\, m_W(W).$$

Similarly, let $x \in \mathcal{M} \setminus \cup_{n\ge 0} \mathcal{S}_{-n}^{\mathbb{H}}$. Then for any $q > 0$ and $A > 0$ there exists an open neighborhood $\mathcal{U}_x^s \subset \mathcal{M}$ of x such that for any stable curve $W \subset \mathcal{U}_x^s$

$$m_W\big(y \in W\colon r^u(y) > A|W|\big) \ge (1-q)\, m_W(W).$$

Next, billiard maps have the following absolute continuity property. Let $W^1, W^2 \subset \mathcal{M}$ be two unstable curves. Denote $W^i_* = \{x \in W^i \colon W^s(x) \cap W^{3-i} \neq \emptyset$ for $i = 1, 2$. The map $\mathbf{h}\colon W^1_* \to W^2_*$ taking every point $x \in W^1_*$ to $\bar{x} = W^s(x) \cap W^2$ ('sliding' it along the stable H-manifold) is called the *holonomy map*. It is absolutely continuous, and its Jacobian (with respect to the Lebesgue measures on W^1 and W^2) is given by

$$J\mathbf{h}(x) = \lim_{n \to \infty} \frac{\mathcal{J}_{W^1}\mathcal{F}^n(x)}{\mathcal{J}_{W^2}\mathcal{F}^n\big(\mathbf{h}(x)\big)}. \tag{2.7}$$

It is uniformly bounded, $J\mathbf{h}(x) \le C$, where $C = C(\mathcal{D}) > 1$ is a constant. Moreover, if we put $\delta = \operatorname{dist}(x, \mathbf{h}(x))$ and denote by γ the angle between the tangent vectors to the curves W^1 and W^2 at the points x and $\mathbf{h}(x)$, respectively, then $J\mathbf{h}(x) \le A^{\gamma + \delta^{1/3}}$ for some constant $A = A(\mathcal{D}) > 1$, see[10]

The Jacobian $J\mathbf{h}(x)$ is a continuous function on W^1_*, but it is not smooth. Even for Anosov and Axiom A systems, the Jacobian of the holonomy map is only Hölder continuous. For billiards, the Hölder continuity may fail, but a similar property holds, it is sometimes called 'dynamically defined Hölder continuity' [26, p. 597]. To describe it, for any $x, y \in \mathcal{M}$ we denote by

$$\mathbf{s}_+(x, y) = \min\{n \ge 0 \colon y \notin \xi^s_n(x)\} \tag{2.8}$$

the 'separation time' (the first time when the images $\mathcal{F}^n(x)$ and $\mathcal{F}^n(y)$ lie in different connected components of the new collision space $\mathcal{M}_{\mathbb{H}}$). Note that $\mathbf{s}_+(x, y) = \infty$ iff $y \in W^s(x)$. Observe that if x and y lie on one unstable curve $W \subset \mathcal{M}$, then

$$|W(x, y)| \le C\Lambda^{-\mathbf{s}_+(x,y)} \tag{2.9}$$

where $C = C(\mathcal{D}) > 0$ is a constant. Then the dynamical Hölder continuity of the Jacobian of the holonomy map is expressed by

$$|\ln J\mathbf{h}(x) - \ln J\mathbf{h}(y)| \le C\boldsymbol{\theta}^{\mathbf{s}_+(x,y)}. \tag{2.10}$$

where $\boldsymbol{\theta} = \Lambda^{-1/6} < 1$ and $C > 0$ are constants, see[10]

Lastly we turn to the so called 'growth lemma.' Let $W \subset \mathcal{M}_{\mathbb{H}}$ be a weakly homogeneous unstable curve and m_W the Lebesgue measure on it. Its image $\mathcal{F}(W) \subset \mathcal{M}_{\mathbb{H}}$ is a finite or countable union of homogeneous unstable curves, which we call the *H-components* of $\mathcal{F}(W)$. Inductively, we define the H-components of $\mathcal{F}^n(W)$ as the collection of the H-components of $\mathcal{F}(W_i)$, $i \ge 1$, where W_i denote all the H-components of $\mathcal{F}^{n-1}(W)$. For every $x \in W$ the point $\mathcal{F}^n(x)$ divides the H-component of $\mathcal{F}^n(W)$ it belongs to into two segments. We denote by $r_n(x)$ the length of the shorter one.

Clearly, $r_n(x)$ is a function on W that characterizes the size of the H-components of $\mathcal{F}^n(W)$. Note that $m_W(r_0(x) < \varepsilon\big) = \min\{2\varepsilon, m_W(W)\}$, where

$m_W(W) = |W|$ is the length of W. The following statements are proved in[9] and [12, Lemma 3.10]:

Lemma 2.2. (Growth Lemma) *There are constants $\hat{\Lambda} \in (1, \Lambda)$, $\vartheta_1 \in (0, 1)$, and $c_1, c_2 > 0$, such that for any sufficiently short unstable curve $W \subset \mathcal{M}$, any $n \geq 0$ and $\varepsilon > 0$*

$$m_W\big(r_n(x) < \varepsilon\big) \leq c_1(\vartheta_1 \hat{\Lambda})^n\, m_W\big(r_0(x) < \varepsilon/\hat{\Lambda}^n\big) + c_2\varepsilon\, m_W(W).$$

Corollary 2.3. *There are constants $\varkappa > 0$ and $c_3 > 0$, such that for all $n \geq \varkappa\big|\ln|W|\big|$ and $\varepsilon > 0$ we have $m_W\big(r_n(x) < \varepsilon\big) \leq c_3\varepsilon\, m_W(W)$,*

3. COUPLING LEMMA

We start with an observation that motivates the use of one-dimensional measures on unstable curves, as proposed by Dolgopyat.

Let us partition the collision space $\mathcal{M}$ into small subdomains (cells) $D_i \subset \mathcal{M}$ and represent a smooth measure μ_0 on $\mathcal{M}$ as a weighted sum of its restrictions to those cells. Now the image of a small domain $D \subset \mathcal{M}$ under the map $\mathcal{F}^n$ gets strongly expanded in the unstable direction, strongly contracted in the stable direction, and possibly cut by singularities into many pieces. Thus, $\mathcal{F}^n(D)$ will soon look like a union of one-dimensional curves, each of which resembles an unstable manifold of the collision map $\mathcal{F}$. Henceforth the measure $\mathcal{F}^n(\mu_0)$ will evolve as a weighted sum of smooth measures on unstable curves.

Accordingly, we define a class of probability measures supported on homogeneous unstable curves. A *standard pair* (W, ν) is a homogeneous unstable curve $W \subset \mathcal{M}$ with a probability measure ν on it, whose density ρ with respect to the Lebesgue measure on W is regular, see below.

The regularity of the density $\rho(x)$ should be comparable to the regularity of the map $\mathcal{F}$, the latter is expressed by two key estimates—distortion bounds and absolute continuity. While distortions are fairly smooth, the Jacobian of the holonomy map is only 'dynamically Hölder continuous' (2.10). Accordingly, we say that a density $\rho(x)$ on a homogeneous unstable curve $W \subset \mathcal{M}$ is regular if

$$|\ln\rho(x) - \ln\rho(y)| \leq C_{\mathrm{r}}\,\boldsymbol{\theta}^{\mathbf{s}_+(x,y)}. \tag{3.1}$$

Here $C_{\mathrm{r}} > 0$ is a sufficiently large constant. Observe that ρ is uniformly bounded:

$$\max_{x\in W}\rho(x)/\min_{x\in W}\rho(x) \leq \text{ const } = e^{C_{\mathrm{r}}}. \tag{3.2}$$

The condition (3.1) will not be altered if we multiply the density $\rho(x)$ by a constant. Therefore, given a standard pair (W, ν), any subcurve $W' \subset W$ with the conditional measure induced by ν on it will make a standard pair. It is easy to see that any unstable H-manifold W^u with the u-SRB measure ν_{W^u} on it makes a standard pair.

The class of standard pairs is invariant under $\mathcal{F}$ in the following sense:

Proposition 3.1. *Let (W, ν) be a standard pair. For each $n \geq 0$, denote by $W_{i,n}$ the H-components of $\mathcal{F}^n(W)$. Then $\mathcal{F}^n(\nu) = \sum_i c_{i,n}\nu_{i,n}$ where $\sum_i c_{i,n} = 1$ and each $(W_{i,n}, \nu_{i,n})$ is a standard pair.*

Proof: By induction, it is enough to prove this for $n = 1$. Let $W_{i,1}$ be an H-component of $\mathcal{F}(W)$ and $x, y \in W_{i,1}$. Denote $x_1 = \mathcal{F}^{-1}(x)$ and $y_1 = \mathcal{F}^{-1}(y)$. Observe that $\mathbf{s}_+(x, y) = \mathbf{s}_+(x_1, y_1) - 1$. Now using (2.1) and (2.9), as well as the relation $\boldsymbol{\theta} = \Lambda^{-1/6}$, gives

$$\begin{aligned} |\ln \rho(x) - \ln \rho(y)| &\leq |\ln \rho_{i,1}(x_1) - \ln \rho_{i,1}(y_1)| \\ &\quad + |\ln \mathcal{J}_W \mathcal{F}^{-1}(x) - \ln \mathcal{J}_W \mathcal{F}^{-1}(y)| \\ &\leq C_{\mathrm{r}} \boldsymbol{\theta}^{\mathbf{s}_+(x_1,y_1)} + C|W_{i,1}|^{1/3} \\ &\leq C_{\mathrm{r}} \boldsymbol{\theta}\, \boldsymbol{\theta}^{\mathbf{s}_+(x,y)} + C' \boldsymbol{\theta}^{\mathbf{s}_+(x,y)} \end{aligned}$$

for some constant $C' = C'(\mathcal{D}) > 0$; here $\rho_{i,1}$ is the density of $\nu_{i,1}$. Thus it is enough to assume that C_{r} is so large that $C_{\mathrm{r}}\boldsymbol{\theta} + C' \leq C_{\mathrm{r}}$.

We see that $\mathcal{F}^n$ transforms the measure ν from a standard pair into a weighted sum of measures on finitely or countably many standard pairs. Motivated by this observation, we introduce even more general families of standard pairs:

A *standard family* is an arbitrary (countable or uncountable) family $\mathcal{G} = \{(W_\alpha, \nu_\alpha)\}$, $\alpha \in \mathfrak{A}$, of standard pairs with a probability factor measure $\lambda_{\mathcal{G}}$ on the index set $\mathfrak{A}$. Such a family induces a probability measure $\mu_{\mathcal{G}}$ on theunion $\cup_\alpha W_\alpha$ (and thus on $\mathcal{M}$) defined by

$$\mu_{\mathcal{G}}(B) = \int \nu_\alpha(B \cap W_\alpha)\, d\lambda_{\mathcal{G}}(\alpha) \qquad \forall B \subset \mathcal{M}.$$

Proposition 3.1 now simply says that $\mathcal{F}^n$ transforms a standard pair into a countable standard family (whose factor measure is defined by the sequence of the coefficients $\{c_{i,n}\}$). Similarly, any standard family $\mathcal{G}$ is mapped by $\mathcal{F}^n$ into another standard family $\mathcal{G}_n = \mathcal{F}^n(\mathcal{G})$. It is easy to see that $\mu_{\mathcal{G}_n} = \mathcal{F}^n(\mu_{\mathcal{G}})$.

It will be important to control the size of curves W_α in a standard family $\mathcal{G} = \{(W_\alpha, \nu_\alpha)\}$. For every $\alpha \in \mathfrak{A}$ and $x \in W_\alpha$, the point x divides the curve W_α into two parts, and we denote by $r_{\mathcal{G}}(x)$ the length of the shorter one (in the

Euclidean metric). This defines a function $r_{\mathcal{G}}$ on[3] $\cup_\alpha W_\alpha$. Now we denote

$$\mathcal{Z}_{\mathcal{G}} = \sup_{\varepsilon>0} \frac{\mu_{\mathcal{G}}(r_{\mathcal{G}} < \varepsilon)}{\varepsilon} = \sup_{\varepsilon>0} \frac{\int \nu_\alpha\big(x \in W_\alpha : r_{\mathcal{G}}(x) < \varepsilon\big)\, d\lambda_{\mathcal{G}}(\alpha)}{\varepsilon}.$$

If $\mathcal{G}$ consists of a single standard pair (W, ν) and ν is is the normalized Lebesgue measure on W, then $\mathcal{Z}_{\mathcal{G}} = 2/|W|$. If ν is an arbitrary regular density, then $\mathcal{Z}_{\mathcal{G}} \sim 1/|W|$, in the sense that $C_1 < \mathcal{Z}_{\mathcal{G}}|W| < C_2$, where $C_1 = C_1(\mathcal{D}) > 0$ and $C_2 = C_2(\mathcal{D}) > 0$ are constants.

Now for an arbitrary standard family $\mathcal{G} = \{(W_\alpha, \nu_\alpha)\}$ we have

$$\mathcal{Z}_{\mathcal{G}} \sim \int \frac{d\lambda_{\mathcal{G}}(\alpha)}{|W_\alpha|} \tag{3.3}$$

(in this formula, either both quantities are finite or both are infinite). We will only consider standard families $\mathcal{G}$ with $\mathcal{Z}_{\mathcal{G}} < \infty$. Next we investigate how the quantity $\mathcal{Z}_{\mathcal{G}_n}$, where $\mathcal{G}_n = \mathcal{F}^n(\mathcal{G})$, changes with n.

Let $\mathcal{G}$ be a standard family consisting of a single standard pair (W, ν). Let $\mathcal{G}_n = \mathcal{F}^n(\mathcal{G})$. The growth lemma 2.2 and (3.2) imply that for all $n \geq 0$ and $\varepsilon > 0$

$$\begin{aligned}\nu\big(r_{\mathcal{G}_n}(\mathcal{F}^n x) < \varepsilon\big) &\leq \text{const}\big[(\vartheta_1 \hat{\Lambda})^n\, \nu\big(r_{\mathcal{G}}(x) < \varepsilon/\hat{\Lambda}^n\big) + \varepsilon\big] \\ &\leq \text{const}\big[\vartheta_1^n \varepsilon/|W| + \varepsilon\big] \end{aligned} \tag{3.4}$$

where the constants depend on the table $\mathcal{D}$ only.

Proposition 3.2. *Let $\mathcal{G} = \{(W_\alpha, \nu_\alpha)\}$, $\alpha \in \mathfrak{A}$, be a standard family and $\mathcal{G}_n = \mathcal{F}^n(\mathcal{G})$. Then for all $n \geq 0$ and $\varepsilon > 0$ we have $\mathcal{Z}_{\mathcal{G}_n} \leq c_1 \vartheta_1^n \mathcal{Z}_{\mathcal{G}} + c_2$ for some constants $c_i = c_i(\mathcal{D}) > 0$, $i = 1, 2$.*

Proof: It is enough to integrate (3.4) with respect to the factor measure $\lambda_{\mathcal{G}}$ and use (3.3).

We see that if $\mathcal{Z}_{\mathcal{G}}$ is very large, the sequence $\mathcal{Z}_{\mathcal{G}_n}$ will decrease exponentially fast until it goes under a certain threshold, say $c_1 + c_2$.

Corollary 3.3. *For all $n \geq \varkappa \ln \mathcal{Z}_{\mathcal{G}}$ we have $\mathcal{Z}_{\mathcal{G}_n} \leq c_3$ for some constants $\varkappa$, $c_3 > 0$.*

The reader should note similarities between Propositions and Corollaries 3.2–3 2.2–2.3

[3] In all our subsequent proofs, the curves W_α of every standard family will be disjoint; however in this general definition we need not assume this: the function $r_{\mathcal{G}}$ is simply defined on every curve W_α separately.

Motivated by these facts, we introduce the notion of a proper family. Let $C_{\mathrm{p}} > 0$ be a sufficiently large constant (the subscript p in C_{p} stands for 'proper'). A standard family $\mathcal{G}$ is said to be *proper* if $\mathcal{Z}_{\mathcal{G}} \leq C_{\mathrm{p}}$. A family consisting of a single standard pair (W, ν) is proper iff $|W| \geq c_0 =$ const > 0, i.e. iff the curve W is not too short (c_0 can be made arbitrarily small by choosing C_{p} sufficiently large). We call such (W, ν) *proper standard pairs*.

The partition ξ^u of $\mathcal{M}$ into maximal unstable H-manifolds with u-SRB measures on them and the factor measure induced by μ makes a (special) standard family $\mathcal{E}$. For this family $\mu_{\mathcal{E}} = \mu$, of course. Note also that $\mathcal{E}$ is mapped by $\mathcal{F}$ into itself. It follows from (2.6) that the family $\mathcal{E}$ is proper.

More generally, let $\{W\}$ be a smooth foliation of $\mathcal{M}$ into unstable curves that stretch from $\varphi = -\pi/2$ to $\varphi = \pi/2$. Dividing them by the homogeneity lines $\mathbb{S}_k$, $|k| \geq k_0$, gives a smooth foliation of $\mathcal{M}$ into homogeneous unstable curves $\{W_\alpha\}$. The measure μ induces smooth conditional measures $\{\nu_\alpha\}$ on $\{W_\alpha\}$ and a factor measure on the index set. It is easy to check that the so defined standard family $\mathcal{G}$ is proper and $\mu_{\mathcal{G}} = \mu$.

Next we turn to the coupling construction. Let (W_1, ν_1) and (W_2, ν_2) be two standard pairs. For a large n, their images $\mathcal{F}^n W_1$ and $\mathcal{F}^n W_2$ consist of many H-components scattered all over $\mathcal{M}$. Some H-components $W' \subset \mathcal{F}^n W_1$ of the image of the first one may lie close to some H-components $W'' \subset \mathcal{F}^n W_2$ of the other image. Then certain points $x' \in W'$ can be joined by stable manifolds with points $x'' \in W''$, so that their further iterations will get closer together exponentially fast. In this way we can 'link' the measures they carry and eventually show that the asymptotic behavior of the two measures $\mathcal{F}^n(\nu_1)$ and $\mathcal{F}^n(\nu_2)$ becomes identical.

This argument may run into an obvious problem, though: the H-components W' and W'' may carry different amount (mass) of the corresponding measures. To resolve this problem we may, so to say, couple a heavy piece with several light ones. This can be done by splitting a heavy piece into several 'thinner' curves, each coupled to a different partner.

To implement this idea, it is convenient to split each original curve W_1 and W_2 into uncountably many 'fibers'. To this end, given a standard pair (W, ν), we consider $\hat{W} := W \times [0, 1]$ and equip $\hat{W}$ with a probability measure $\hat{\nu}$ defined by

$$d\hat{\nu}(x, t) = d\nu(x)\, dt = \rho(x)\, dx\, dt \tag{3.5}$$

where $\rho(x)$ is the density of ν and $0 \leq t \leq 1$. We call $\hat{W}$ a *rectangle* with *base* W. The map $\mathcal{F}^n$ can be naturally defined on $\hat{W}$ by $\mathcal{F}^n(x, t) = (\mathcal{F}^n x, t)$ and any function f initially defined on W can be also extended to $\hat{W}$ by $f(x, t) = f(x)$.

Next, recall that $\mathcal{F}^n(W_1)$ and $\mathcal{F}^n(W_2)$ are, generally, countable standard families. Thus, in the above construction we may start with two standard families, rather than two standard pairs, and couple the images of measures initially defined on the two families. In that case we need to split *every* unstable curve W in *each* original family into uncountable many 'fibers' by constructing a rectangle over W.

Given a standard family $\mathcal{G} = (W_\alpha, \nu_\alpha)$, $\alpha \in \mathfrak{A}$, with a factor measure $\lambda_{\mathcal{G}}$, we denote by $\hat{\mathcal{G}} = (\hat{W}_\alpha, \hat{\nu}_\alpha)$ the family of the corresponding rectangles, with $\hat{W}_\alpha$ being the rectangle with base W_α, equip $\hat{\mathcal{G}}$ with the same factor measure $\lambda_{\mathcal{G}}$, and denote by $\hat{\mu}_{\mathcal{G}}$ the induced measure on the union $\cup_\alpha \hat{W}_\alpha$.

The following lemma is the key instrument of the coupling method

Lemma 3.4. (Coupling Lemma) *Let $\mathcal{G} = (W_\alpha, \nu_\alpha)$, $\alpha \in \mathfrak{A}$, and $\mathcal{E} = (W_\beta, \nu_\beta)$, $\beta \in \mathfrak{B}$, be two proper standard families.*[4] *Then there exist a bijection (a coupling map) $\Theta\colon \cup_\alpha \hat{W}_\alpha \to \cup_\beta \hat{W}_\beta$ that preserves measure, i.e. $\Theta(\hat{\mu}_{\mathcal{G}}) = \hat{\mu}_{\mathcal{E}}$, and a (coupling time) function $\Upsilon : \cup_\alpha \hat{W}_\alpha \to \mathbb{N}$ such that*

A. Let $(x, t) \in \hat{W}_\alpha$, $\alpha \in \mathfrak{A}$, and $\Theta(x, t) = (y, s) \in \hat{W}_\beta$, $\beta \in \mathfrak{B}$. Denote $m = \Upsilon(x, t) \in \mathbb{N}$. Then the points $\mathcal{F}^m(x)$ and $\mathcal{F}^m(y)$ lie on the same stable H-manifold $W^s \subset \mathcal{M}$.

B. There is a uniform exponential tail bound on the function Υ: we have $\hat{\mu}_{\mathcal{G}_1}\big(\Upsilon > n\big) \le C_\Upsilon \vartheta_\Upsilon^n$ for some constants $C_\Upsilon = C_\Upsilon(\mathcal{D}) > 0$ and $\vartheta_\Upsilon = \vartheta_\Upsilon(\mathcal{D}) < 1$.

The proof of Coupling Lemma will be given in Appendix. Next we introduce a class of appropriate functions on $\mathcal{M}$ (observables). They will be characterized by dynamically defined Hölder continuity.

Similar to the (future) separation time $\mathbf{s}_+(x, y)$ defined by (2.8), we introduce the *past separation time*:

$$\mathbf{s}_-(x, y) = \min\{n \ge 0\colon y \notin \xi_n^u(x)\} \tag{3.6}$$

(the first time in the past when the preimages $\mathcal{F}^{-n}(x)$ and $\mathcal{F}^{-n}(y)$ lie in different connected components of the 'new' collision space $\mathcal{M}_{\mathbb{H}}$). If x and y lie on one stable curve $W \subset \mathcal{M}$, then

$$\operatorname{dist}(x, y) \le C\Lambda^{-\mathbf{s}_-(x,y)} \tag{3.7}$$

for some constant $C = C(\mathcal{D}) > 0$. We say that a function $f\colon \mathcal{M} \to \mathbb{R}$ is *dynamically Hölder continuous* if there are $\vartheta_f \in (0, 1)$ and $K_f > 0$ such that for any x and y lying on one unstable curve

$$|f(x) - f(y)| \le K_f \vartheta_f^{\mathbf{s}_+(x,y)} \tag{3.8}$$

and for any x and y lying on one stable curve

$$|f(x) - f(y)| \le K_f \vartheta_f^{\mathbf{s}_-(x,y)} \tag{3.9}$$

[4] Here $\mathcal{E}$ may be the proper standard family defined above. However, it is easy to see that we can just as well use two arbitrary proper standard families$\hat{\mu}$.

We denote the space of such functions by $\mathcal{H}$. The class of dynamically Hölder continuous functions contains the class of (ordinary) Hölder continuous functions; the latter are characterized by

$$|f(x) - f(y)| \le C_f[\mathrm{dist}(x, y)]^{\alpha_f} \qquad \forall x, y \in \mathcal{M}, \tag{3.10}$$

here $\alpha_f \in (0, 1]$ is the Hölder exponent and the minimal $C_f > 0$ satisfying (3.10) is the Hölder norm of f. Indeed, any Hölder continuous function $f : \mathcal{M} \to \mathbb{R}$ is dynamically Hölder continuous with $\vartheta_f = \Lambda^{-\alpha_f}$.

Furthermore, suppose that f is piecewise Hölder continuous, i.e. there are $n_1, n_2 \ge 0$ such that f is Hölder continuous on every connected component of the set $\mathcal{M} \setminus (\mathcal{S}^{\mathbb{H}}_{n_1} \cup \mathcal{S}^{\mathbb{H}}_{-n_2})$, i.e. (3.10) holds, with the same C_f and α_f, whenever x and y belong in the same component. Then again f is dynamically Hölder continuous.

It is easy to verify by direct inspection that in billiards with finite horizon the return time function $\tau(x)$ is Hölder continuous on the connected components of the set $\mathcal{M} \setminus \mathcal{S}_1$, hence it is dynamically Hölder continuous.

Remark. Occasionally we will deal with functions satisfying only one of the conditions (3.8) and (3.9). We denote the space of functions satisfying (3.8) by $\mathcal{H}^+$, and those satisfying (3.9) by $\mathcal{H}^-$.

4. EQUIDISTRIBUTION AND DECAY OF CORRELATIONS

First we use Coupling Lemma to show that, given a proper standard family $\mathcal{G}$, the images $\mathcal{F}^n(\mu_{\mathcal{G}})$ of its measure $\mu_{\mathcal{G}}$ weakly converge, as $n \to \infty$, to the $\mathcal{F}$-invariant measure μ; furthermore, in a certain sense the speed of convergence is exponential. We call this property *equidistribution*.

Theorem 4.1. (Equidistribution) *Let $\mathcal{G}$ be a proper standard family. For any dynamically Hölder continuous function $f \in \mathcal{H}$ and $n \ge 0$*

$$\left| \int_{\mathcal{M}} f \circ \mathcal{F}^n \, d\mu_{\mathcal{G}} - \int_{\mathcal{M}} f \, d\mu \right| \le B_f \theta_f^n \tag{4.1}$$

where $B_f = 2C_\Upsilon \big(K_f + \|f\|_\infty\big)$ and

$$\theta_f = \big[\max\{\vartheta_\Upsilon, \vartheta_f\}\big]^{1/2} < 1. \tag{4.2}$$

Proof: Recall that there is a proper $\mathcal{F}$-invariant family $\mathcal{E}$ such that $\mu = \mu_{\mathcal{E}}$. The coupling lemma 3.4 gives us a coupling map Θ between the families $\mathcal{G}$ and $\mathcal{E}$ and the corresponding coupling time function Υ. Then

$$\Delta := \int_{\mathcal{M}} f \circ \mathcal{F}^n \, d\mu_{\mathcal{G}} - \int_{\mathcal{M}} f \circ \mathcal{F}^n \, d\mu_{\mathcal{E}}$$

$$= \int_{\hat{\mathcal{G}}} f(\mathcal{F}^n(x,t))\, d\hat{\mu}_{\mathcal{G}} - \int_{\hat{\mathcal{E}}} f(\mathcal{F}^n(y,s))\, d\hat{\mu}_{\mathcal{E}}$$

$$= \int_{\hat{\mathcal{G}}} \big[f(\mathcal{F}^n(x,t)) - f(\mathcal{F}^n(\Theta(x,t)))\big]\, d\hat{\mu}_{\mathcal{G}}. \tag{4.3}$$

Note that if $\Theta(x,t) = (y,s)$ and $m\colon = \Upsilon(x,t) \le n$, then $\mathbf{s}_-(\mathcal{F}^n x, \mathcal{F}^n y) > n - m$, hence by the clause A of Lemma 3.4

$$\big| f(\mathcal{F}^n(x,t) - f(\mathcal{F}^n(\Theta(x,t)))\big| \le K_f \vartheta_f^{n-m}. \tag{4.4}$$

Now the last integral in (4.3) can be decomposed as

$$\int_{\hat{\mathcal{G}}} [\cdots] = \int_{\Upsilon \le n/2} [\cdots] + \int_{\Upsilon > n/2} [\cdots] = I + II. \tag{4.5}$$

Observe that (4.4) implies $|I| \le K_f \vartheta_f^{n/2}$. Also, the clause B implies that $|II| \le 2\,\|f\|_\infty C_\Upsilon \vartheta_\Upsilon^{n/2}$.

Theorem 4.1 can be extended to 'multiple' observables, i.e. observations made at multiple moments of time. Let $f_0, f_1, \dots, f_k \in \mathcal{H}$ be dynamically Hölder continuous functions with the same $\vartheta_f = \vartheta_{f_i}$, the same $K_f = K_{f_i}$, and the same $\|f\|_\infty = \|f_i\|_\infty$, $0 \le i \le k$. (For example, we can take $f_0 = f_1 = \cdots = f_k = f$.) Consider the product $\tilde{f} = f_0 \cdot (f_1 \circ \mathcal{F}) \cdot (f_2 \circ \mathcal{F}^2) \cdots (f_k \circ \mathcal{F}^k)$.

Theorem 4.2. *Let $\mathcal{G}$ be a proper standard family. Then for $n \ge 0$*

$$\left| \int_{\mathcal{M}} \tilde{f} \circ \mathcal{F}^n \, d\mu_{\mathcal{G}} - \int_{\mathcal{M}} \tilde{f}\, d\mu \right| \le B_{\tilde{f}}\, \theta_f^n \tag{4.6}$$

where $\theta_f < 1$ is as in (4.2) and $B_{\tilde{f}} = 2C_\Upsilon \|f\|_\infty^k (1 - \vartheta_f)^{-1}\big(K_f + \|f\|_\infty\big)$.

Proof: The argument is almost identical to the proof of Theorem 4.1, except (4.4) must be replaced by

$$\begin{aligned} \big|\tilde{f}(\mathcal{F}^n(x,t)) - \tilde{f}(\mathcal{F}^n(\Theta(x,t)))\big| &\le K_f \|f\|_\infty^k \big(\vartheta_f^{n-m} + \cdots + \vartheta_f^{n-m+k}\big) \\ &\le K_f \|f\|_\infty^k (1-\vartheta_f)^{-1} \vartheta_f^{n-m}. \end{aligned}$$

Remark. We used consecutive time moments $0, 1, \dots, k$ for simplicity. The statement (and the proof) will not change if we consider any increasing sequence of time moments $0 < t_1 < t_2 < \cdots < t_k$.

Remark. In Theorems 4.1 and 4.2 we assumed that the initial standard family $\mathcal{G}$ was proper. If it is not, then we will have to wait until its image $\mathcal{F}^m \mathcal{G}$ becomes proper for some $m \ge 1$ and then apply these theorems with n replaced by $n - m$. If $\mathcal{Z}_{\mathcal{G}} < \infty$, then the waiting time is $m = \varkappa \ln \mathcal{Z}_{\mathcal{G}}$ iterations of $\mathcal{F}$, according to

Corollary 3.2. If $\mathcal{G}$ consists of a single standard pair (W, ν), the waiting time is $m = \varkappa \left|\ln |W|\right|$.

REMARK. In Theorems 4.1 and 4.2 it is enough to assume that $f \in \mathcal{H}^-$ (respectively, $f_0, \ldots, f_k \in \mathcal{H}^-$), cf. Remark in the end of the previous section.

Next we derive an exponential bound on correlations for dynamically Hölder continuous observables. Similar bounds were obtained earlier[26,97], but ours is sharper; its advantage will be apparent in the next section. For brevity, we use notation $\langle f \rangle = \int_{\mathcal{M}} f\, d\mu$.

Theorem 4.3. (Exponential decay of correlations) *For every pair of dynamically Hölder continuous functions $f, g \in \mathcal{H}$ and $n \geq 0$*

$$\left|\langle f \cdot (g \circ \mathcal{F}^n)\rangle - \langle f \rangle \langle g \rangle\right| \leq B_{f,g}\, \theta_{f,g}^n \tag{4.7}$$

where

$$\theta_{f,g} = \left[\max\left\{\vartheta_\Upsilon, \vartheta_f, \vartheta_g, e^{-1/\varkappa}\right\}\right]^{1/4} < 1, \tag{4.8}$$

where $\varkappa > 0$ is the constant of Theorem 2.2,

$$B_{f,g} = C_0\left(K_f \|g\|_\infty + K_g \|f\|_\infty + \|f\|_\infty \|g\|_\infty\right), \tag{4.9}$$

and $C_0 = C_0(\mathcal{D}) > 0$ is a constant.

Proof: We again use the proper standard family $\mathcal{E} = \{W_\alpha, \nu_\alpha\}$, $\alpha \in \mathfrak{A}$, such that $\mu_{\mathcal{E}} = \mu$. We can write

$$\langle f \cdot (g \circ \mathcal{F}^n)\rangle = \int \left(f \circ \mathcal{F}^{-n/4}\right)\left(g \circ \mathcal{F}^{3n/4}\right) d\mu_{\mathcal{E}}.$$

Let $\bar{f}$ denote the conditional expectation of $f \circ \mathcal{F}^{-n/4}$ on the unstable H-manifolds W_α with respect to the u-SRB measure ν_α, i.e.

$$\bar{f}(x) = \int_{W_\alpha} f \circ \mathcal{F}^{-n/4}\, d\nu_\alpha \qquad \forall x \in W_\alpha \quad \forall \alpha \in \mathfrak{A}.$$

Clearly the function $f \circ \mathcal{F}^{-n/4}$ is almost constant on each H-manifold W_α. Precisely, due to the dynamical Hölder continuity of f

$$\sup_{W_\alpha} f \circ \mathcal{F}^{-n/4} - \inf_{W_\alpha} f \circ \mathcal{F}^{-n/4} \leq K_f \vartheta_f^{n/4}, \tag{4.10}$$

hence $\sup_{x \in \mathcal{M}} |\bar{f}(x) - f \circ \mathcal{F}^{-n/4}(x)| \leq K_f \vartheta_f^{n/4}$. Note also that $\langle f \rangle = \langle \bar{f} \rangle$, thus

$$\begin{aligned} \Delta :&= \langle f \cdot (g \circ \mathcal{F}^n)\rangle - \langle f \rangle \langle g \rangle \\ &= \langle \bar{f} \cdot (g \circ \mathcal{F}^{3n/4})\rangle - \langle \bar{f} \rangle \langle g \rangle + \delta_1, \end{aligned}$$

where $|\delta_1| \leq K_f \|g\|_\infty \vartheta_f^{n/4}$. Since the function $\bar{f}$ is constant on every H-manifold W_α, we denote its value by $\bar{f}_\alpha$ and use Theorem 4.1 and the second remark after Theorem 4.2 to obtain

$$\int_{W_\alpha} \bar{f} \cdot (g \circ \mathcal{F}^{3n/4})\, d\nu_\alpha = \bar{f}_\alpha \langle g \rangle + \delta_\alpha \tag{4.11}$$

where

$$|\delta_\alpha| \leq 2 \min\Big\{ \|f\|_\infty \|g\|_\infty, \\ \|f\|_\infty C_\Upsilon \big(K_g + \|g\|_\infty\big) \theta_g^{3n/4 - \varkappa |\ln|W_\alpha||} \Big\}$$

and $\theta_g = \big[\max\{\vartheta_\Upsilon, \vartheta_g\}\big]^{1/2}$. Integrating (4.11) with respect to the factor measure $\lambda_\mathcal{E}$ of the family $\mathcal{E}$ gives

$$\langle \bar{f} \cdot (g \circ \mathcal{F}^{3n/4}) \rangle = \langle \bar{f} \rangle \langle g \rangle + \int_{\mathfrak{A}} \delta_\alpha \, d\lambda_\mathcal{E}(\alpha).$$

It remains to estimate the last term here. Since $\mathcal{E}$ is a proper family, we have $\mu\big(\cup W_\alpha \colon |W_\alpha| < e^{-\frac{n}{4\varkappa}}\big) \leq C_\mathrm{p} e^{-\frac{n}{4\varkappa}}$. For H-manifolds W_α satisfying $|W_\alpha| \geq e^{-\frac{n}{4\varkappa}}$ we have $3n/4 - \varkappa \big|\ln|W_\alpha|\big| \geq n/2$. Therefore

$$\left| \int \delta_\alpha \, d\lambda_\mathcal{E}(\alpha) \right| \leq 2 C_\mathrm{p} \|f\|_\infty \|g\|_\infty e^{-\frac{n}{4\varkappa}} \\ + 2 \|f\|_\infty C_\Upsilon \big(K_g + \|g\|_\infty\big) \theta_g^{n/2}.$$

Theorem 4.3 is now proved.

In this theorem, it is enough to assume that $f \in \mathcal{H}^+$ and $g \in \mathcal{H}^-$, cf. our earlier remarks. This observation leads to an important corollary. Suppose f is constant on every unstable H-manifold W^u and g is constant on every stable H-manifold W^s. We can also redefine unstable and stable cones so that they will degenerate to lines tangent to the unstable and stable manifolds, respectively, see Remark on cones in Section 2. Then we can assume that $K_f = K_g = 0$ and $\theta_f = \theta_g = 0$, and therefore

$$\big| \langle f \cdot (g \circ \mathcal{F}^n) \rangle - \langle f \rangle \langle g \rangle \big| \leq C_0 \|f\|_\infty \|g\|_\infty \theta_0^n \tag{4.12}$$

for all $n \geq 0$, where $\theta_0 = \big[\max\big\{\vartheta_\Upsilon, e^{-1/\varkappa}\big\}\big]^{1/4} < 1$. This follows from (4.7)–(4.9).

Corollary 4.4. *Let $A, B \subset \mathcal{M}$ be two measurable sets such that $A = \cup W^u$ is the union of* some *unstable H-manifolds and $B = \cup W^s$ be the union of* some *stable H-manifolds (here we mean* maximal *H-manifolds, cf. Section 2). Then*

$$\big| \mu\big(A \cap \mathcal{F}^{-n}(B)\big) - \mu(A)\mu(B) \big| \leq C_0 \theta_0^n. \tag{4.13}$$

Theorem 4.3 can be extended to ‘multiple’ correlations, i.e. correlations between observations made at multiple moments of time. This sharply improve earlier results[(11,14)]. Let $f_0, f_1, \dots, f_r \in \mathcal{H}$ and $g_0, g_1, \dots, g_k \in \mathcal{H}$ be two sets of dynamically Hölder continuous functions. We suppose f’s have identical parameters $\vartheta_f = \vartheta_{f_i}$, $K_f = K_{f_i}$, and $\|f\|_\infty = \|f_i\|_\infty$ for all $0 \le i \le r$. Similarly, let g’s have identical parameters $\vartheta_g = \vartheta_{g_i}$, $K_g = K_{g_i}$, and $\|g\|_\infty = \|g_i\|_\infty$ for all $0 \le i \le k$. Consider two products $\tilde f = f_0 \cdot (f_1 \circ \mathcal{F}^{-1}) \cdot (f_2 \circ \mathcal{F}^{-2}) \cdots (f_r \circ \mathcal{F}^{-r})$ and $\tilde g = g_0 \cdot (g_1 \circ \mathcal{F}) \cdot (g_2 \circ \mathcal{F}^2) \cdots (g_k \circ \mathcal{F}^k)$.

Theorem 4.5. (Exponential decay of multiple correlations) *For all $n > 0$*

$$\left|\langle \tilde f \cdot (\tilde g \circ \mathcal{F}^n)\rangle - \langle \tilde f\rangle\langle \tilde g\rangle\right| \le B_{\tilde f,\tilde g}\, \theta_{f,g}^{|n|} \tag{4.14}$$

where $\theta_{f,g}$ is as in (4.8) and

$$B_{\tilde f,\tilde g} = C_0 \|f\|_\infty^r \|g\|_\infty^k \left[\frac{K_f \|g\|_\infty}{1 - \vartheta_f} + \frac{K_g \|f\|_\infty}{1 - \vartheta_g} + \|f\|_\infty \|g\|_\infty\right] \tag{4.15}$$

and $C_0 = C_0(\mathcal{D}) > 0$ is a constant.

Proof: The argument is almost identical to the proof of Theorem 4.3, with a few modifications. First, we note obvious bounds: $\|\tilde f\|_\infty \le \|f\|_\infty^{r+1}$ and $\|\tilde g\|_\infty \le \|g\|_\infty^{k+1}$. Second, (4.10) is replaced with

$$\begin{aligned}\sup_{W_\alpha} \tilde f \circ \mathcal{F}^{-n/4} - \inf_{W_\alpha} \tilde f \circ \mathcal{F}^{-n/4} &\le K_f \|f\|_\infty^r \big(\vartheta_f^{n/4} + \cdots + \vartheta_f^{n/4+r}\big)\\ &\le K_f \|f\|_\infty^r (1 - \vartheta_f)^{-1} \vartheta_f^{n/4}.\end{aligned}$$

Lastly, Theorem 4.2 must be used instead of 4.1.

Remark. We used consecutive time moments $-r, \dots, -1, 0$ and $n, n+1, \dots, n+k$ for simplicity. The statement (and the proof) will be the same for any two increasing sequences of time moments $t_{-r} < \cdots < t_{-1} < 0$ and $n < t_1 < \cdots < t_k$ separated by a ‘time gap’ of length n.

5. LIMIT THEOREMS

First we recall relevant definitions. Given a measure preserving map $F\colon M \to M$ and a function $f\colon M \to \mathbb{R}$, we denote its partial sums by $S_n = \sum_{i=0}^{n-1} f \circ F^i$. Assume that $\langle f\rangle = 0$. We say that f satisfies the Central Limit Theorem (CLT) if

$$\lim_{n\to\infty} \mu\left\{\frac{S_n}{\sqrt{n}} \le z\right\} = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{z} e^{-\frac{s^2}{2\sigma^2}}\, ds \tag{5.1}$$

for all $-\infty < z < \infty$. Here $\sigma = \sigma_f \geq 0$ is related to correlations by

$$\sigma_f^2 = \sum_{n=-\infty}^{\infty} \langle f \cdot (f \circ F^n) \rangle = \langle f^2 \rangle + 2 \sum_{n=1}^{\infty} \langle f \cdot (f \circ F^n) \rangle. \tag{5.2}$$

In addition, the degenerate case $\sigma_f^2 = 0$ occurs if and only if the function f is cohomologous to zero, i.e. $f = g - g \circ F$ for some $g \in L^2(M)$. This is a rather general fact, see(17) and [16 Theorem 18.2.2]

Next, let f satisfy the central limit theorem and $\sigma_f > 0$. For $N \geq 1$ and $x \in M$ consider continuous function $W_N(s; x)$ of $s \in [0, 1]$ defined by

$$W_N\left(\frac{n}{N}; x\right) = \frac{S_n(x)}{\sigma_f \sqrt{N}}$$

at rational points $s = n/N$ and by linear interpolation in between. The invariant measure μ and the family $\{W_N(s; x), x \in M\}$ induce a probability measure on the space of continuous functions on$^{[0,1]}$.

We say that f satisfies *Weak Invariance Principle* (WIP) if the above measure weakly converges, as $N \to \infty$, to the Wiener measure. We say that f satisfies *Almost Sure Invariance Principle* (ASIP) if there is a standard Wiener process (a Brownian motion) $B(s; x)$ on M with respect to the measure μ so that for some $\lambda > 0$

$$|W_N(s; x) - B(s; x)| = \mathcal{O}(N^{-\lambda})$$

for μ-almost all $x \in M$. The invariance principle thus asserts that S_n, after a proper rescaling of space and time, converges to the Wiener process.

In this section we prove that every dynamically Hölder continuous observable $f \in \mathcal{H}$ satisfies the CLT, WIP, and ASIP, along with a few other limit laws. We use a general theorem established by Philipp and Stout(19), which we present here in a form adapted to our needs.

Let ξ_0 be a finite or countable partition of M. Denote $\xi_n = F^n(\xi_0)$ and $\xi_m^n = \xi_m \vee \cdots \vee \xi_n$ for all $m < n$ (we allow $m = -\infty$ and $n = \infty$). Each partition ξ_m^n has its σ-algebra $\mathfrak{F}_m^n$ consisting of measurable ξ_m^n-sets. We assume that $\xi_{-\infty}^{\infty}$ is the maximal partition, i.e. $\mathfrak{F}_{-\infty}^{\infty}$ is the full σ-algebra of all measurable sets in M.

Let $f \colon M \to \mathbb{R}$ be a measurable function. For $m \geq 1$, denote by $\bar{f}_m = E\big(f | \mathfrak{F}_{-m}^m\big)$ the conditional expectation of f on the σ-algebra $\mathfrak{F}_{-m}^m$, i.e.

$$\bar{f}_m(x) = \frac{1}{\mu(B)} \int_B f \, d\mu, \qquad \forall x \in B \quad \forall B \in \xi_{-m}^m.$$

The following theorem is proved in(19), see Theorem 7.1 and Remark on p. 81 in that book.

Theorem 5.1. [19] *Suppose that there exist constants* $0 < \delta < 2$ *and* $C > 0$ *such that*

$$\langle |f(x)|^{2+\delta} \rangle < \infty \tag{5.3}$$

and for all $m \geq 1$

$$\langle |f(x) - \bar{f}_m|^{2+\delta} \rangle \leq C m^{-(2+7/\delta)}. \tag{5.4}$$

Moreover, suppose that $\sigma_f^2 > 0$ *and*

$$\operatorname{Var} S_n = n\sigma_f^2 + \mathcal{O}\big(n^{1-\delta/30}\big) \tag{5.5}$$

as $n \to \infty$. *Finally, suppose that for all* $n \geq 1$

$$|\mu(A \cap B) - \mu(A)\mu(B)| \leq C n^{-168(1+2/\delta)} \tag{5.6}$$

for all $A \in \mathfrak{F}_{-\infty}^0$ *and* $B \in \mathfrak{F}_n^\infty$. *Then the sequence* $\mathcal{X}_n = f \circ F^n$ *satisfies the CLT, WIP, and ASIP.*

The ASIP is the strongest claim here - it actually implies the WIP, which in turn implies the CLT. The ASIP also implies many other limit laws, we only mention some and refer the reader to[19, Section 1] for more.

Corollary 5.2. (Integral Tests) *Let* $\phi(t)$ *be a positive nondecreasing real-valued function. Then*

$$\mu\Big(\frac{S_n - n\langle f \rangle}{\sqrt{n}} > \phi(n) \ \text{ infinitely often}\Big) = 0 \text{ or } 1$$

according as the integral $\int_1^\infty \frac{\phi(t)}{t}\, e^{-\phi^2(t)/2}\, dt$ *converges or diverges. Furthermore, put* $M_n = \max_{1 \leq i \leq n} |S_i - i\langle f \rangle|$. *Then*

$$\mu\big(M_n/\sqrt{n} < \phi^{-1}(n) \ \text{ infinitely often}\big) = 0 \text{ or } 1$$

according as the integral $\int_1^\infty \frac{\phi^2(t)}{t}\, e^{-8\pi^{-2}\phi^2(t)}\, dt$ *converges or diverges.*

Corollary 5.3. (Law of Iterated Logarithm) *For* μ*-almost every point* $x \in M$

$$\limsup_{n \to \infty} \frac{S_n - n\langle f \rangle}{\sqrt{2n\sigma_f^2 \log\log n}} = 1.$$

We now return to the collision map $\mathcal{F}\colon \mathcal{M} \to \mathcal{M}$ of a Sinai billiard. To apply the above theorem It remains to prove the following:

Lemma 5.4. *The collision map $\mathcal{F}\colon \mathcal{M} \to \mathcal{M}$ for dispersing billiards and dynamically Hölder continuous observables $f \in \mathcal{H}$ such that $\sigma_f^2 > 0$ satisfy all the conditions of Theorem 5.1.*

Proof: Observe that (5.3) holds for any $\delta \in (0, 2)$ because $\|f\|_\infty < \infty$. Next, the exponential decay of correlations easily implies an even stronger property than (5.5): $\operatorname{Var} S_n = n\sigma_f^2 + \mathcal{O}(1)$, see [8, Section 3]. To prove (5.4) and (5.6) we need to construct the partition ξ_0.

Let ξ_0 be the partition of $\mathcal{M}$ into the connected components of the set $\mathcal{M} \setminus \mathcal{S}^{\mathbb{H}}$. That is, the elements of ξ_0 are domains on which the map $\mathcal{F}\colon \mathcal{M}_{\mathbb{H}} \to \mathcal{M}_{\mathbb{H}}$ is smooth. Recall that every element of ξ_0 is a curvilinear polygon, bounded by stable curves (and horizontal lines belonging to $\partial\mathcal{M}$), and the interior angles of these polygons do not exceed π.

Then $\xi_1 = \mathcal{F}(\xi_0)$ is the partition of $\mathcal{M}$ into the connected components of the set $\mathcal{M} \setminus \mathcal{S}_{-1}^{\mathbb{H}}$, so its elements are similar curvilinear polygons bounded by unstable curves (and horizontal lines belonging to $\partial\mathcal{M}$). It is easy to see that for any $A \in \xi_0$ and $B \in \xi_1$ the intersection $A \cap B$ is a curvilinear polygon, i.e. the elements of $\xi_0^1 = \xi_0 \vee \xi_1$ are the connected components of the set $\mathcal{M} \setminus (\mathcal{S}^{\mathbb{H}} \cup \mathcal{S}_{-1}^{\mathbb{H}})$.

By induction, one can verify directly that for any $p, q > 0$ the elements of the partition ξ_{-p}^q are the connected components of the set $\mathcal{M} \setminus (\mathcal{S}_{p+1}^{\mathbb{H}} \cup \mathcal{S}_{-q}^{\mathbb{H}})$. Therefore, the elements of the partition ξ_{-m}^m are the connected components of the set $\mathcal{M} \setminus (\mathcal{S}_{m+1}^{\mathbb{H}} \cup \mathcal{S}_{-m}^{\mathbb{H}})$. Also, they are curvilinear polygons bounded by stable and unstable curves, and their interior angles do not exceed π.

Let $A \in \xi_{-m}^m$. Obviously, for any two points $x, y \in A$ there is a point $z \in A$ such that x and z belong to one unstable curve and y and z belong to one stable curve. Furthermore, $\mathbf{s}_+(x, z) > m$ and $\mathbf{s}_-(y, z) > m$, thus $|f(x) - f(y)| \leq 2K_f\vartheta_f^m$. It follows that $\langle |f(x) - \bar{f}_m|^{2+\delta}\rangle \leq \text{const}\, \vartheta_f^{(2+\delta)m}$. This implies (5.4).

Lastly we prove the main hypothesis (5.6). Observe that ξ_1^∞ coincides with the partition ξ^u of $\mathcal{M}$ into (maximal) unstable H-manifolds introduced in Section 2. Similarly, $\xi_{-\infty}^0$ coincides with the partition ξ^s of $\mathcal{M}$ into (maximal) stable H-manifolds. Hence any set $A \in \mathfrak{F}_{-\infty}^0$ is a union of some stable H-manifolds. Similarly, for any set $B \in \mathfrak{F}_n^\infty$ its preimage $\mathcal{F}^{-n+1}(B) \in \mathfrak{F}_1^\infty$ will be a union of some unstable H-manifolds. Now Corollary 4.3 easily implies that $|\mu(A \cap B) - \mu(A)\mu(B)| \leq C_0\theta_0^{n-1}$. This proves (5.6).

We remark that the CLT and WIP can also be proved more directly, see(7,.8).

APPENDIX

Here we prove the coupling Lemma 3.4. Our argument is quite lengthy but fairly transparent and completely dynamical.

Any closed region $\mathfrak{Q} \subset \mathcal{M}$ bounded by two unstable H-manifolds and two stable H-manifolds will be called a *solid rectangle*. Its boundary consists of four smooth curves, which we naturally call the *u-sides* and *s-sides*. If an unstable H-manifold $W^u \subset \mathcal{M}$ crosses both s-sides of $\mathfrak{Q}$, we say that it *fully crosses* the solid rectangle $\mathfrak{Q}$. Similar notions apply to stable H-manifolds.

Given a solid rectangle $\mathfrak{Q} \subset \mathcal{M}$, denote by $\mathfrak{R} = \mathfrak{R}(\mathfrak{Q})$ the set of points $x \in \mathfrak{Q}$ such that *both* H-manifolds $W^u(x)$ and $W^s(x)$ fully cross $\mathfrak{Q}$. The set $\mathfrak{R}$ is a closed nowhere dense Cantor-like subset of $\mathfrak{Q}$ that has a natural direct-product structure. More generally, a closed subset $\mathfrak{R} \subset \mathcal{M}$ is called a *Cantor rectangle* (or just *rectangle*) if for any $x, y \in \mathfrak{R}$ the intersection $W^u(x) \cap W^s(y)$ consists of one point that belongs to $\mathfrak{R}$.

Let $\mathfrak{R}$ be a rectangle and $z \in \mathfrak{R}$. The set $W^u(z) \cap \mathfrak{R}$ is closed and lies on an increasing curve, thus it has two extreme points, call them x_1 and x_2. Similarly, let y_1 and y_2 denote the two extreme points of the set $W^s(z) \cap \mathfrak{R}$. Then the two stable H-manifolds $W^s(x_i)$, $i = 1, 2$, and two unstable H-manifolds $W^u(y_i)$, $i = 1, 2$ enclose a solid rectangle $\mathfrak{Q}$ containing $\mathfrak{R}$. We denote it by $\mathfrak{Q}(\mathfrak{R})$ and call it the *hull* of the rectangle $\mathfrak{R}$.

Let $\mathfrak{R}$ be a rectangle. A subset $\mathfrak{R}_1 \subset \mathfrak{R}$ is called a *u-subrectangle* if $W^u(x) \cap \mathfrak{R} = W^u(x) \cap \mathfrak{R}_1$ for any point $x \in \mathfrak{R}_1$. Similarly, $\mathfrak{R}_2 \subset \mathfrak{R}$ is called an *s-subrectangle* if $W^s(x) \cap \mathfrak{R} = W^s(x) \cap \mathfrak{R}_1$ for any point $x \in \mathfrak{R}_2$.

The image $\mathcal{F}^n(\mathfrak{R})$ of a rectangle $\mathfrak{R}$ for any $n \in \mathbb{Z}$ is a finite or countable union of rectangles $\{\mathfrak{R}_i\}$. For $n > 0$, their preimages $\mathcal{F}^{-n}(\mathfrak{R}_i)$ are s-subrectangles in $\mathfrak{R}$. For $n < 0$, they are u-subrectangles in $\mathfrak{R}$.

Due to absolute continuity, $\mu(\mathfrak{R}) > 0$ if and only if for any (and hence, for every) point $z \in \mathfrak{R}$ we have[5] $|W^u(z) \cap \mathfrak{R}| > 0$ and $|W^s(z) \cap \mathfrak{R}| > 0$. We will only deal with rectangles of positive measure. We call

$$\rho^u(\mathfrak{R}) = \inf_{x \in \mathfrak{R}} \frac{|W^u(x) \cap \mathfrak{R}|}{|W^u(x) \cap \mathfrak{Q}(\mathfrak{R})|}$$

the (minimal) *u-density* of $\mathfrak{R}$. Similarly the (minimal) *s-density* $\rho^s(\mathfrak{R})$ is defined and we call $\rho(\mathfrak{R}) = \min\{\rho^u(\mathfrak{R}), \rho^s(\mathfrak{R})\}$ the (minimal) *density* of the rectangle $\mathfrak{R}$. Due to the compactness of $\mathfrak{R}$ and the continuity of the above ratio (in x), we have $\mu(\mathfrak{R}) > 0$ if and only if $\rho(\mathfrak{R}) > 0$.

Proposition A.1. *For any point $x \in \mathcal{M}$ that has non-vanishing H-manifolds $W^u(x)$ and $W^s(x)$, there is a closed rectangle $\mathfrak{R} \ni x$ of positive measure. Moreover, for any $\delta > 0$, we can find a rectangle $\mathfrak{R} \ni x$ with density $\rho(\mathfrak{R}) > 1 - \delta$ and such that the point x divides the curves $W^u(x) \cap \mathfrak{Q}(\mathfrak{R})$ and $W^s(x) \cap \mathfrak{Q}(\mathfrak{R})$ in the ratio between $0.5 - \delta$ and $0.5 + \delta$, i.e. x is almost a geometric center of $\mathfrak{Q}(\mathfrak{R})$.*

[5] We denote by $|W|$ the length of a curve W, thus for any subset $B \subset W$ the expression $|B|$ means the one-dimensional Lebesgue measure of B.

Proof: Since $W^u(x)$ and $W^s(x)$ exist, we have $x \in \mathcal{M} \setminus \left(\cup_{n=-\infty}^{\infty} \mathcal{S}_n^{\mathbb{H}}\right)$. Now the claim easily follows from Sinai's Fundamental Theorem 2.1.

Next we construct a special rectangle $\mathfrak{R}_*$ whose stable manifolds will be used to 'connect' (or 'couple') points of $\mathcal{F}^n(\cup_\alpha W_\alpha)$ with those of $\mathcal{F}^n(\cup_\beta W_\beta)$. The rectangle $\mathfrak{R}_*$, like a magnet, will 'attract' the H-components of the images of our proper standard families.

Let $W^u \subset \mathcal{M}$ be an unstable H-manifold. Recall that $r_n(x)$ denotes the distance from the point $\mathcal{F}^n(x)$ to the nearest endpoint of the H-component of $\mathcal{F}^n(W^u)$ that contains the point $\mathcal{F}^n(x)$. Also, recall that $r^s(x)$ denotes the distance, measured along the (maximal) stable H-manifold $W^s(x)$, from x to the nearest endpoint of $W^s(x)$. It is a rather standard fact of hyperbolic dynamics that

$$r^s(x) \geq \min_{n \geq 0} \tilde{C}^{-1} \Lambda^n \, r_n(x), \tag{A.1}$$

where $\Lambda > 1$ is the minimal expansion factor and $\tilde{C} = \tilde{C}(\mathcal{D}) > 0$ is a constant.

Now let $\tilde{W} \subset W^u$ be subcurve (to be chosen later). Given $\kappa > 0$, we now put

$$\tilde{W}_\kappa := \tilde{W} \setminus \cup_{n \geq 0} \{x \in \tilde{W} : r_n(x) < \tilde{C} \Lambda^{-n} \kappa\}. \tag{A.2}$$

Note that the subset $\tilde{W}_\kappa \subset \tilde{W}$ is closed. Due to (A.1), we have $r^s(x) \geq \kappa$ for every $x \in \tilde{W}_\kappa$. We denote by $\mathfrak{S}_\kappa(\tilde{W}) = \{W^s(x) : x \in \tilde{W}_\kappa\}$ the set of the corresponding stable H-manifolds (all of them extend the distance $\geq \kappa$ from $\tilde{W}$ on both sides).

Since $\tilde{W}_\kappa$ is a closed set, it has two extreme points, x_1 and x_2, on the curve $\tilde{W}$, which correspond to two extreme H-manifolds $W^s(x_1)$ and $W^s(x_2)$ in the family $\mathfrak{S}_\kappa(\tilde{W})$. We say that an unstable curve W *fully crosses* our family $\mathfrak{S}_\kappa(\tilde{W})$, if it crosses *all* the H-manifolds $W^s \in \mathfrak{S}_\kappa(\tilde{W})$.

Denote by $\mathcal{G}_\kappa^u(\tilde{W})$ the family of all unstable H-manifolds $W^u \subset \mathcal{M}$ that fully cross the family $\mathfrak{S}_\kappa(\tilde{W})$. Lastly let

$$\mathfrak{R}_\kappa(\tilde{W}) = \cup_{W^s \in \mathfrak{S}_\kappa(\tilde{W})} \cup_{W^u \in \mathcal{G}_\kappa^u(\tilde{W})} W^s \cap W^u$$

denote the rectangle made by our two families of stable and unstable H-manifolds.

Proposition A.2. *For any $\delta > 0$ there are a subcurve $\tilde{W} \subset W^u$ and a $\kappa > 0$ such that the rectangle $\mathfrak{R}_* = \mathfrak{R}_\kappa(\tilde{W})$ has density $\rho(\mathfrak{R}_*) > 1 - \delta$.*

Proof: First let $\tilde{W} = W^s$. It follows from (2.6) that $\left|\tilde{W} \setminus \cup_{\kappa > 0} \tilde{W}_\kappa\right| = 0$. Choose $\kappa > 0$ such that $|\tilde{W}_\kappa| > 0$ and pick a Lebesgue density point $z \in \tilde{W}_\kappa$ on the curve $\tilde{W}$. Now reduce $\tilde{W}$ so that it becomes a small neighborhood of z. The rest of the proof is the same as in Proposition A.1.

From now on we choose a small $\delta > 0$ and fix the corresponding $\tilde{W} \subset W^u$ and $\kappa > 0$ and the special rectangle $\mathfrak{R}_* = \mathfrak{R}_\kappa(\tilde{W})$ constructed in the above proposition. We denote $\mathfrak{S} = \mathfrak{S}_\kappa(\tilde{W})$, for brevity, and slightly abusing notation we also denote by $\mathfrak{S}$ the union of all the H-manifolds $W^s \in \mathfrak{S}$. Note that $\tilde{W}_\kappa = \tilde{W} \cap \mathfrak{S}$. For any unstable curve W that fully crosses $\mathfrak{S}$ we set $W_\kappa := W \cap \mathfrak{S}$.

Next, for any standard pair (W, ν) and any $n \geq 0$ denote by $W_{n,i}$ all the H-components of $\mathcal{F}^n(W)$ that fully cross $\mathfrak{S}$ and put

$$W_{n,*} = \cup_i \mathcal{F}^{-n}(W_{n,i} \cap \mathfrak{S}). \tag{A.3}$$

The following is proved in [7, Theorem 3.13]

Proposition A.3. *There are constants $n_1 \geq 1$ and $d_1 > 0$ such that for any proper standard pair and any $n \geq n_1$ we have $\nu(W_{n,*}) \geq d_1$.*

This easily extends to proper standard families, with obvious notation:

Corollary A.4. *There are constants $n_0 \geq 1$ and $d_0 > 0$ such that for any proper standard family $\mathcal{G} = \{(W_\alpha, \nu_\alpha)\}$ and any $n \geq n_0$ we have $\mu_{\mathcal{G}}\big(\cup_\alpha W_{\alpha,n,*}\big) \geq d_0$.*

Remark. In Proposition A.3 we assumed that the standard pair (W, ν) was proper. If it is not, then we have to wait until its image $\mathcal{F}^m(W)$ becomes a proper standard family for some $m \geq 1$ and then apply Corollary A.4, with n_0 replaced by $n_0 + m$. Recall that the waiting time here is $m = \varkappa\ \big|\ln|W|\big|$.

Let $\mathcal{P} = (W, \nu)$ be a standard pair such that W fully crosses the family $\mathfrak{S}$ (the magnet) constructed above. Then $W_\kappa = W \cap \mathfrak{S}$ is a closed nowhere dense set on the curve W, and its complement $W \setminus W_\kappa$ consists of infinitely many intervals; we call them *gaps in* W_κ. These gaps naturally correspond to the connected components of the set $\tilde{W} \setminus \tilde{W}_\kappa$ (gaps in $\tilde{W}_\kappa$) which are created by the points $x \in \tilde{W}$ satisfying $r_n(x) < \tilde{C}\Lambda^{-n}\kappa$, in accordance with (A.2).

Let $\tilde{V} \subset \tilde{W} \setminus \tilde{W}_\kappa$ be an interval. We call

$$n = \min\{i \geq 1 : r_i(x) < \tilde{C}\Lambda^{-i}\kappa \text{ for some } x \in \tilde{V}\}$$

the *rank* of $\tilde{V}$. Clearly every gap in $\tilde{W}_k$ has a rank. If rank $\tilde{V} = n$, then $\mathcal{F}^{n-1}(\tilde{V})$ is a curve of length $\geq C\Lambda^{-n}$ for some constant $C = C(\mathcal{D}) > 0$. Indeed, consider the H-component of $\mathcal{F}^{n-1}(\tilde{W})$ containing $\mathcal{F}^{n-1}(\tilde{V})$; it intersects the $(C\Lambda^{-n})$-neighborhood of the singularity set $\mathcal{S}^{\mathbb{H}}$ for some constant $C > 0$; then it has to cross this neighborhood completely (for if it terminates somewhere inside that neighborhood, then it must have been torn by the singularities some time earlier).

Now every gap $V \subset W \setminus W_\kappa$ corresponds to a gap $\tilde{V} \subset \tilde{W} \setminus \tilde{W}_\kappa$ that has some rank $n \geq 1$; in this case we also say that V itself has rank n. The endpoints of V are linked to those of $\tilde{V}$ by stable H-manifolds $W^s_{\mathbb{H}} \in \mathfrak{S}$, hence their images $\mathcal{F}^{n-1}(V)$ and $\mathcal{F}^{n-1}(\tilde{V})$ must be exponentially close to each other (the distance between them is $\ll \Lambda^{-n}$). Therefore $\left|\mathcal{F}^{n-1}(V)\right| \geq \frac{1}{2}C\Lambda^{-n}$. It follows that the set $\mathcal{F}^{n(2+\ \varkappa\ \ln\Lambda)}(V)$ will be a proper standard family, in accordance with Remark after Corollary A.4.

Accordingly, we define *recovery time* function $\mathbf{r}_{\mathcal{P}}(x)$ on $W \setminus W_\kappa$ by setting $\mathbf{r}_{\mathcal{P}}(x) = [n(2 + \varkappa \ln \Lambda)]$, where n is the rank of the gap $V \subset W \setminus W_\kappa$ containing the point x (note that the function $\mathbf{r}_{\mathcal{P}}(x)$ is now constant on every gap).

Corollary 2.2 implies that for all $n \geq 1$

$$\nu\big(x \in W \setminus W_\kappa \colon \mathbf{r}_{\mathcal{P}}(x) > n\big) \leq \text{const}\, \Lambda^{-n}\, \nu(W \setminus W_\kappa). \tag{A.4}$$

Next, let $\mathbf{s}_{\mathcal{P}}(x)$ be another function on $W \setminus W_\kappa$ (with values in $\mathbb{N}$) such that

$$\mathbf{s}_{\mathcal{P}}(x) \equiv \text{const} \qquad \text{on every gap } V \subset W \setminus W_\kappa \tag{A.5}$$

and

$$\mathbf{s}_{\mathcal{P}}(x) \geq \mathbf{r}_{\mathcal{P}}(x) + n_0. \tag{A.6}$$

Since both functions $\mathbf{r}_{\mathcal{P}}$ and $\mathbf{s}_{\mathcal{P}}$ are constant on each gap V, we will occasionally denote their values by $\mathbf{r}_{\mathcal{P}}(V)$ and $\mathbf{s}_{\mathcal{P}}(V)$, respectively.

Now applying Corollary A.4 to every gap $V \subset W \setminus W_\kappa$ (more precisely, to its image $\mathcal{F}^{\mathbf{r}_{\mathcal{P}}(V)}(V)$, cf. Remark after Corollary A.4) gives

$$\nu\big(V_{\mathbf{s}_{\mathcal{P}}(V),*}\big) \geq d_0\, \nu(V), \tag{A.7}$$

in the notation of (A.3). In other words at time $n = \mathbf{s}_{\mathcal{P}}(V)$ the d_0-fraction of the image $\mathcal{F}^n(V)$ will be ‘on the magnet’. At this time the image $\mathcal{F}^n(V)$ may be ‘stopped’ and part of it (which intersects the magnet $\mathfrak{S}$) may be coupled with the corresponding image of another proper standard family (this will be done below). We call $\mathbf{s}_{\mathcal{P}}$ the *stopping time* function.

Observe that the stopping time $\mathbf{s}_{\mathcal{P}}$ is not yet completely specified by (A.5) and (A.6); that is the purpose of the next proposition.

Proposition A.5. *We can define the stopping time function* $\mathbf{s}_{\mathcal{P}}(x)$ *on* $W \setminus W_\kappa$ *(see the remark below) so that for all* $n \in \mathbb{N}$

$$\frac{\nu\big(x \in W \setminus W_\kappa \colon \mathbf{s}_{\mathcal{P}}(x) = n\big)}{\nu(W \setminus W_\kappa)} = q_n, \tag{A.8}$$

where $\{q_n\}$ *is a sequence satisfying*

$$\sum q_n = 1 \quad \text{and} \quad q_n < \text{const}\, \theta^n \tag{A.9}$$

for some $\theta \in (\Lambda^{-1}, 1)$. *Furthermore, the sequence* $\{q_n\}$ *is independent of* $\mathcal{P}$, *i.e. it is the same for all standard pairs* $\mathcal{P} = (W, \nu)$ *such that* W *fully crosses the family* $\mathfrak{S}$.

Proof: Due to (A.4), it is easy to define $\mathbf{s}_{\mathcal{P}}$ so that for all $n > 0$

$$\frac{\nu\big(x \in W \setminus W_\kappa \colon \mathbf{s}_{\mathcal{P}}(x) = n\big)}{\nu(W \setminus W_\kappa)} \leq \text{const}\, \theta^n.$$

We still have a considerable flexibility in defining the function $\mathbf{s}_{\mathcal{P}}$, and we want to adjust it so that it will satisfy (A.8) with a sequence $\{q_n\}$ being independent of $\mathcal{P}$. To this end we split every gap V into an uncountable family of 'thinner' curves with the help of rectangles described in Section 3. Precisely, we replace each gap V with a rectangle $V \times [0, 1]$. Then we can divide the latter into subrectangles $V \times I_j$, where $I_j \subset [0, 1]$ are some subintervals, and define $\mathbf{s}_{\mathcal{P}}$ so that it takes a different value on each subrectangle I_j. The sizes of the subintervals $I_j \subset [0, 1]$ must be selected to ensure (A.8), as well as (A.9).

Since our $\mathbf{s}_{\mathcal{P}}$ is now constant on every subrectangle $V \times I_j$, the latter can be then collapsed to a curve V_j, which geometrically coincides with the gap V, but carries a measure different from ν; precisely, it carries the measure ν multiplied by $|I_j|$. In the end we will have a countable family of curves $\{V_j\}$ and the function $\mathbf{s}_{\mathcal{P}}$ will be constant on each of them, as desired.

Remark. Observe that in the proof of Proposition A.5 we had to split some of the components of $W \subset W_\kappa$ into finitely or countably many (geometrically identical) curves, each having a different weight, and define the stopping time function $\mathbf{s}_{\mathcal{P}}$ separately of each curves. This small correction is needed to make the proposition precise.

We now turn to the construction of the coupling map $\Theta\colon \cup_\alpha \hat{W}_\alpha \to \cup_\beta \hat{W}_\beta$ for Lemma 3.4. This will be done recursively, in an algorithm-like manner. The first two steps of our construction will be described in detail, and then it will be clear how it proceeds.

Recall that we are given two proper standard families $\mathcal{G} = (W_\alpha, \nu_\alpha)$, $\alpha \in \mathfrak{A}$, and $\mathcal{E} = (W_\beta, \nu_\beta)$, $\beta \in \mathfrak{B}$, with the corresponding measures $\mu_{\mathcal{G}}$ and $\mu_{\mathcal{E}}$. We denote by $\hat{\mathcal{G}} = (\hat{W}_\alpha, \hat{\nu}_\alpha)$ and $\hat{\mathcal{E}} = (\hat{W}_\beta, \hat{\nu}_\beta)$ the respective families of rectangles, and we have two probability measures $\hat{\mu}_{\mathcal{G}}$ and $\hat{\mu}_{\mathcal{E}}$ on the unions $\cup_\alpha \hat{W}_\alpha$ and $\cup_\beta \hat{W}_\beta$, respectively.

We define the first stopping time function $\mathbf{s}_0$ on the unions $\cup_\alpha \hat{W}_\alpha$ and $\cup_\beta \hat{W}_\beta$ to be constant $\mathbf{s}_0(x, t) \equiv n_0$. At time $\mathbf{s}_0 = n_0$ some of the H-components of the images $\mathcal{F}^{\mathbf{s}_0}\big(\cup_\alpha W_\alpha\big)$ and $\mathcal{F}^{\mathbf{s}_0}\big(\cup_\beta W_\beta\big)$ will fully cross the magnet $\mathfrak{S}$; and the total measure of the respective intersections with $\mathfrak{S}$ will be $\geq d_0$ due to Corollary A.4.

For every H-component $W_{\alpha,\mathbf{s}_0,i}$ of $\mathcal{F}^{\mathbf{s}_0}(W_\alpha)$ that fully crosses $\mathfrak{S}$ we consider the corresponding rectangle $\hat{W}_{\alpha,\mathbf{s}_0,i} = W_{\alpha,\mathbf{s}_0,i} \times [0, 1]$. We now split off a subrectangle $W_{\alpha,\mathbf{s}_0,i} \times [0, \boldsymbol{\tau}_{\alpha,i}]$ with some $0 < \boldsymbol{\tau}_{\alpha,i} \leq 0.5$ so that

$$\hat{\mu}_{\mathcal{G}}(\cup_\alpha \tilde{W}_{\alpha,1}) = d \colon= d_0/2, \tag{A.10}$$

where

$$\tilde{W}_{\alpha,1} = \{(x, t) \in \hat{W}_\alpha \colon \mathcal{F}^{\mathbf{s}_0}(x) \in W_{\alpha,\mathbf{s}_0,i} \cap \mathfrak{S} \\ \& \; t \in [0, \boldsymbol{\tau}_{\alpha,i}] \text{ for some } i\big\}$$

This can be done easily due to Corollary A.4: if $\mu_{\mathcal{G}}(\cup_\alpha W_{\alpha,\mathbf{s}_0,*}) = d_0$ in Corollary A.4, we simply set $\boldsymbol{\tau}_{\alpha,i} = 0.5$ for all α and i; if the inequality in Corollary A.4 is strict, we have "too much of a good thing", then we lower some of $\boldsymbol{\tau}_{\alpha,i}$'s to make (A.10) exact.

We will put tildas over W's that denote subsets of $\cup_\alpha \hat{W}_\alpha$ on which we are currently defining the coupling map Θ. At the first step of our construction, Θ must take points $(x,t) \in \cup_\alpha \tilde{W}_{\alpha,1}$ to points $(y,s) \in \cup_\beta \hat{W}_\beta$ such that $\mathcal{F}^{\mathbf{s}_0}(x)$ and $\mathcal{F}^{\mathbf{s}_0}(y)$ lie on the same stable H-manifold $W^s \in \mathfrak{S}$. It also must preserve measure (i.e., take $\hat{\mu}_{\mathcal{G}}$ to $\hat{\mu}_{\mathcal{E}}$). To correctly define Θ on the set $\cup_\alpha \tilde{W}_{\alpha,1}$ we will first describe its image, which we will denote by $\cup_\beta \tilde{W}_{\beta,1}$ (here the index 1 refers, of course, to the *first* step of our construction).

One may be tempted to define $\tilde{W}_{\beta,1}$'s in the same way as we defined $\tilde{W}_{\alpha,1}$'s above. In that case the sets $\cup_\beta \tilde{W}_{\beta,1}$ and $\cup_\alpha \tilde{W}_{\alpha,1}$ would have the same overall measure ($= d$), and their $\mathcal{F}^{\mathbf{s}_0}$-images would lie on the same stable H-manifolds $W^s \in \mathfrak{S}$, but this may not suffice, as such a map may not preserve measure. Indeed, for some $W^s \in \mathfrak{S}$ the intersections $W^s \cap \mathcal{F}^{\mathbf{s}_0}(\cup_\alpha \tilde{W}_{\alpha,1})$ and $W^s \cap \mathcal{F}^{\mathbf{s}_0}(\cup_\beta \tilde{W}_{\beta,1})$ may carry different 'amount' of measures $\hat{\mu}_{\mathcal{G}}$ and $\hat{\mu}_{\mathcal{E}}$, respectively. There are two possible reasons for this 'mismatch': (i) the densities of our measures may vary along our H-components and (ii) the Jacobian of the holonomy map may also vary and differ from one.

To deal with the possible 'mismatch', we first assume, without loss of generality, that the diameter of the 'special rectangle' $\mathfrak{R}_*$ is very small; so that the corresponding oscillations of the densities are at least very small (say, the ratio of the densities at different points on the same H-component is between 0.99 and 1.01), and the Jacobian of the holonomy map takes values in a narrow interval around one, say, in [0.99, 1.01].

We now define the set $\cup_\beta \tilde{W}_{\beta,1}$ as follows. For every H-component $W_{\beta,\mathbf{s}_0,j} \subset \mathcal{F}^{\mathbf{s}_0}(W_\beta)$ that fully crosses the magnet $\mathfrak{S}$ we will construct a function $\boldsymbol{\tau}_{\beta,j}(y)$ on $W_{\beta,\mathbf{s}_0,j} \cap \mathfrak{S}$, with values in the interval [0, 0.6] (this function will later be extended to the entire curve $W_{\beta,\mathbf{s}_0,j}$), and then put

$$\begin{aligned}\tilde{W}_{\beta,1} = \{(y,t) \in \hat{W}_\beta \colon \mathcal{F}^{\mathbf{s}_0}(y) \in W_{\beta,\mathbf{s}_0,j} \cap \mathfrak{S} \\ \& \; t \in [0, \boldsymbol{\tau}_{\beta,j}(\mathcal{F}^{\mathbf{s}_0} y)] \;\text{ for some } j\big\}.\end{aligned}$$

The functions $\boldsymbol{\tau}_{\beta,j}$ can be constructed so that for every stable H-manifold $W^s \in \mathfrak{S}$ the intersections $W^s \cap \mathcal{F}^{\mathbf{s}_0}(\cup_\alpha \tilde{W}_{\alpha,1})$ and $W^s \cap \mathcal{F}^{\mathbf{s}_0}(\cup_\beta \tilde{W}_{\beta,1})$ carry the same 'amount' of measures $\hat{\mu}_{\mathcal{G}}$ and $\hat{\mu}_{\mathcal{E}}$, respectively. This is the reason why we need to give some room to the functions $\boldsymbol{\tau}_{\beta,j}$ so that their values can be adjusted accordingly, this is why we allow them to go up to 0.6 (as compared to $\boldsymbol{\tau}_{\alpha,i}$ that took values ≤ 0.5).

Also, since the densities of measures for standard pairs are dynamically Hölder continuous (3.1), and so is the Jacobian of the holonomy map, it follows that the functions $\boldsymbol{\tau}_{\beta,j}(y)$ will be dynamically Hölder continuous as well, i.e. they will satisfy

$$|\ln \boldsymbol{\tau}_{\beta,j}(y) - \ln \boldsymbol{\tau}_{\beta,j}(z)| \le C_0 \boldsymbol{\theta}^{s_+(y,z)} \tag{A.11}$$

for some constant $C_0 > 0$.

We now naturally define the coupling map $\Theta\colon \cup_\alpha \tilde{W}_{\alpha,1} \to \cup_\beta \tilde{W}_{\beta,1}$ that preserves measure and couples points whose $\mathcal{F}^{\mathbf{s}_0}$-images lie on the same stable manifold of the $\mathfrak{S}$ family. We note that

$$\hat{\mu}_{\mathcal{G}}\big(\cup_\alpha \tilde{W}_{\alpha,1}\big) = \hat{\mu}_{\mathcal{E}}\big(\cup_\beta \tilde{W}_{\beta,1}\big) = d, \tag{A.12}$$

where the constant $d = d_0/2$ was introduced in (A.10). Lastly we set the coupling time function $\Upsilon(x,t) = \mathbf{s}_0$ on $\cup_\alpha \tilde{W}_{\alpha,1}$. This concludes the first step of our construction of the coupling map Θ.

Before we move on to the second step, we need to 'inventory' the remaining (uncoupled) parts of the families $\mathcal{G}$ and $\mathcal{E}$ and represent their images at time $\mathbf{s}_0$ by unions of some rectangles. To this end we first define a constant function $\boldsymbol{\tau}_{\alpha,i}(x)$ on every H-component $W_{\alpha,\mathbf{s}_0,i}$ of $\mathcal{F}^{\mathbf{s}_0}(W_\alpha)$ that fully crosses $\mathfrak{S}$ so that $\boldsymbol{\tau}_{\alpha,i}(x) \equiv \boldsymbol{\tau}_{\alpha,i}$, where $\boldsymbol{\tau}_{\alpha,i}$ is the constant chosen earlier (before equation (A.10)).

On the contrary, the function $\boldsymbol{\tau}_{\beta,j}(y)$ defined earlier on the Cantor-like subset $W_{\beta,\mathbf{s}_0,j} \cap \mathfrak{S} \subset W_{\beta,\mathbf{s}_0,j}$ is *not* constant. We now extend it to the entire curve $W_{\beta,\mathbf{s}_0,j}$ by linear interpolation (we make it linear on every gap $V \subset W_{\beta,\mathbf{s}_0,j} \setminus \mathfrak{S}$ and overall continuous). The graph of $\boldsymbol{\tau}_{\beta,j}$ divides the rectangle $W_{\beta,\mathbf{s}_0,j} \times [0,1]$ into two parts ('subrectangles', each has one irregular side, see Fig. 2). It is easy to verify that the function $\boldsymbol{\tau}_{\beta,j}(y)$, after its extension to $W_{\beta,\mathbf{s}_0,j}$, is still dynamically Hölder continuous in the sense of (A.11).

Now the 'uncoupled' sets $\cup_\alpha \mathcal{F}^{\mathbf{s}_0}\big(\hat{W}_\alpha \setminus \tilde{W}_{\alpha,1}\big)$ and $\cup_\beta \mathcal{F}^{\mathbf{s}_0}\big(\hat{W}_\beta \setminus \tilde{W}_{\beta,1}\big)$ consist of connected components of three types.

First, there are rectangles corresponding to the H-components of $\mathcal{F}^{\mathbf{s}_0}(W_\alpha)$ and $\mathcal{F}^{\mathbf{s}_0}(W_\beta)$ that do not fully cross the magnet $\mathfrak{S}$, we did not modify them in any way.

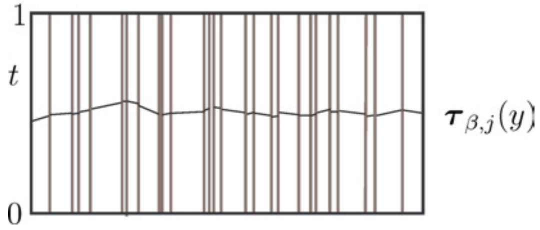

Fig. 2. The partition of a rectangle over an H-component $W_{\beta,\mathbf{s}_0,j}$: the irregular line in the middle is the graph of the function $\boldsymbol{\tau}_{\beta,j}(y)$; it separates the 'upper subrectangle' (of the second type) from lower trapezoids (of the third type).

Second, the 'upper subrectangles'

$$\{(x,t)\colon x \in W_{\alpha,\mathbf{s}_0,i} \;\&\; t \in [\boldsymbol{\tau}_{\alpha,i}(x), 1]\} \tag{A.13}$$

and similar regions

$$\{(y,t)\colon y \in W_{\beta,\mathbf{s}_0,j} \;\&\; t \in [\boldsymbol{\tau}_{\beta,j}(x), 1]\} \tag{A.14}$$

(the latter are not genuine subrectangles, they have one 'jagged' side as shown on Fig. 2). All of the regions (A.13)–(A.14) have sufficiently long bases (longer than the size of the special rectangle $\mathfrak{R}_*$ in its unstable direction).

Third, the 'lower subrectangles'

$$\{(x,t)\colon x \in V \;\&\; t \in [0, \boldsymbol{\tau}_{\alpha,i}(x)]\}$$

constructed over gaps $V \subset W_{\alpha,\mathbf{s}_0,i} \setminus \mathfrak{S}$ and similar regions

$$\{(y,t)\colon y \in V' \;\&\; t \in [0, \boldsymbol{\tau}_{\beta,j}(y)]\}$$

constructed over gaps $V' \subset W_{\beta,\mathbf{s}_0,j} \setminus \mathfrak{S}$ (the latter are trapezoids as shown Fig. 5).

Next we 'rectify' the irregular sides of the rectangles of the second and third type by a simple algorithm: it consists of stretching of each vertical (i.e. parallel to the t axis) fiber inside every rectangle accordingly. More precisely, given a 'rectangle' $\hat{W}_1 = \{(x,t)\colon x \in W \;\&\; t \in [0, \boldsymbol{\tau}(x)]\}$, where W is an unstable curve and $\boldsymbol{\tau}(x)\colon W \to [0,1]$ is a continuous function, equipped with a probability measure $d\hat{\nu}(x,t) = \rho(x)\,dx\,dt$, we transform the interval $[0, \boldsymbol{\tau}(x)]$ onto the unit interval $[0,1]$ linearly at every point $x \in W$, and thus obtain a full-height rectangle $\hat{W} = W \times [0,1]$ with measure

$$d\hat{\nu}_1(x,t) = \rho_1(x)\,dx\,dt, \qquad \rho_1(x) = \boldsymbol{\tau}(x)\,\rho(x). \tag{A.15}$$

Recall that the 'ceiling function' $\boldsymbol{\tau}(x)$ is dynamically Hölder continuous for the rectangles of the second and third type, and so is every regular density $\rho(x)$ according to (3.1). Hence the new density $\rho_1(x)$ defined by (A.15) will be also dynamically Hölder continuous, precisely

$$|\ln \rho_1(x) - \ln \rho_1(y)| \le (C_0 + C_{\mathrm{r}})\,\boldsymbol{\theta}^{\mathbf{s}_+(x,y)}. \tag{A.16}$$

Of course the constant $C = C_0 + C_{\mathrm{r}}$ is larger than C_{r} in (3.1), so the density $\rho_1(x)$ is not necessarily regular (yet). But its images will smooth out (a similar phenomenon was exploited in the proof of Proposition 3.1) and become regular in $m_0 \ge 1$ iterations of $\mathcal{F}$, where $m_0 \ge 1$ is a constant (this follows from distortion bounds). Moreover, by making C_{r} larger, if necessary, we can even ensure that $m_0 = 1$, i.e. the densities regularize right away.

Thus the remaining (uncoupled) sets

$$\cup_\alpha \mathcal{F}^{\mathbf{s}_0}\big(\hat{W}_\alpha \setminus \tilde{W}_{\alpha,1}\big) \quad \text{and} \quad \cup_\beta\, \mathcal{F}^{\mathbf{s}_0}\big(\hat{W}_\beta \setminus \tilde{W}_{\beta,1}\big) \tag{A.17}$$

are unions of rectangles of the full (unit) height. We denote the families of those rectangles by $\hat{\mathcal{G}}_1 = \{\hat{W}_{\alpha,1}\}$ and $\hat{\mathcal{E}}_1 = \{\hat{W}_{\beta,1}\}$, respectively. (Note that the $\mathcal{F}^{\mathbf{s}_0}$-image of an 'old rectangle' $\hat{W}_\alpha$ may contain countably many 'new' rectangles $\hat{W}_{\alpha,1}$, thus we have to reindex our families, so the new indices α and β may not correspond to the old ones used for the original families $\mathcal{G}$ and $\mathcal{E}$, but this will cause no harm.)

So we get two new families $\hat{\mathcal{G}}_1$ and $\hat{\mathcal{E}}_1$, each carries a measure induced by the $\mathcal{F}^{\mathbf{s}_0}$-image of the original measure ($\hat{\mu}_{\mathcal{G}}$ or $\hat{\mu}_{\mathcal{E}}$). Conditioning the induced measures on the new families $\hat{\mathcal{G}}_1$ and $\hat{\mathcal{E}}_1$ gives two probability measures on them, we call them $\hat{\mu}_{\hat{\mathcal{G}}_1}$ and $\hat{\mu}_{\hat{\mathcal{E}}_1}$, respectively. The densities of the new measures $\hat{\mu}_{\hat{\mathcal{G}}_1}$ and $\hat{\mu}_{\hat{\mathcal{E}}_1}$ may not be regular, but their images become regular in just $\leq m_0$ iterations on $\mathcal{F}$ (or even in just one step, see above), so we disregard this slight inconvenience.

The main complication is that the new families $\hat{\mathcal{G}}_1$ and $\hat{\mathcal{E}}_1$ may not be proper, because they contain myriad of arbitrarily small rectangles of the third type created in the gaps of the magnet $\mathfrak{S}$. Of course, if we condition the measures $\hat{\mu}_{\hat{\mathcal{G}}_1}$ and $\hat{\mu}_{\hat{\mathcal{E}}_1}$ onto the union of rectangles of the first and second type, then the so reduced families (albeit not necessarily proper either) will obviously recover and become proper standard families in just a few iterations of $\mathcal{F}$, we leave the verification of this simple fact to the reader; so we may assume that the rectangles of the first and second type make a proper standard family already.

However, on the rectangles of the third type, the recovery time may vary greatly. We define the stopping time function $\mathbf{s}_1(x,t)$ on the rectangles of the third type as described in Proposition 5. In particular, the function $\mathbf{s}_1$ takes values in $\mathbb{N}$, is constant on every rectangle, and it corresponds to the time when the image of the rectangle becomes a proper standard family plus an extra n_0 iterations of $\mathcal{F}$.

We need also to define the stopping time function $\mathbf{s}_1$ on the rectangles of the first and the second types, so that it takes values in $\mathbb{N}$, is constant on every rectangle, and its overall distribution on *all* rectangles matches the one described in Proposition 5, i.e.

$$\hat{\mu}_{\hat{\mathcal{G}}_1}\big(\cup_\alpha \hat{W}_{\alpha,1} \colon \mathbf{s}_1 = n\big) = q_n \qquad \forall n \in \mathbb{N} \tag{A.18}$$

and

$$\hat{\mu}_{\hat{\mathcal{E}}_1}\big(\cup_\beta \hat{W}_{\beta,1} \colon \mathbf{s}_1 = n\big) = q_n \qquad \forall n \in \mathbb{N}. \tag{A.19}$$

with the same sequence $\{q_n\}$ as in (A.8)–(A.9).

In order to define such a function $\mathbf{s}_1$ and ensure (A.18) and (A.19) we may need to split some rectangles $\hat{W}_{\alpha,1}$ and $\hat{W}_{\beta,1}$ of the first and second type into 'thinner' subrectangles, as we did in the proof of Proposition A.5, and define $\mathbf{s}_1$ separately on every subrectangle. Since the family of rectangles of the first and second type is proper already, this task is much simpler than the proof of Proposition 5, so we leave details to the reader.

Now we are in a position very similar to the one we were earlier. For every rectangle $\hat{W}_{\alpha,1}$, the set $\mathcal{F}^{\mathbf{s}_1}(\hat{W}_{\alpha,1})$ with the induced measure will be a proper family, it will contain H-components fully crossing the magnet $\mathfrak{S}$, and their intersections with $\mathfrak{S}$ will have a relative measure $\geq d_0$ due to Corollary A.4:

$$\frac{\hat{\mu}_{\hat{\mathcal{G}}_1}(\hat{W}_{\alpha,1,\mathbf{s}_1,*})}{\hat{\mu}_{\hat{\mathcal{G}}_1}(\hat{W}_{\alpha,1})} \geq d_0 \quad \text{and} \quad \frac{\hat{\mu}_{\hat{\mathcal{E}}_1}(\hat{W}_{\beta,1,\mathbf{s}_1,*})}{\hat{\mu}_{\hat{\mathcal{E}}_1}(\hat{W}_{\beta,1})} \geq d_0 \tag{A.20}$$

for every α and β, in the notation of (A.3) and Proposition A.0. We note that $\mathbf{s}_1$ is constant on every rectangle $\hat{W}_{\alpha,1}$ and $\hat{W}_{\beta,1}$, so these notations make sense.

Of course, the value of $\mathbf{s}_1$ in (A.20) depends on α (or β). In what follows, we will group rectangles $\hat{W}_{\alpha,1}$ and $\hat{W}_{\beta,1}$ on which the function $\mathbf{s}_1$ takes the same value. In particular, we have

$$\hat{\mu}_{\hat{\mathcal{G}}_1}\big(\cup_\alpha \hat{W}_{\alpha,1,\mathbf{s}_1,*}: \mathbf{s}_1 = n\big) \geq d_0\, \hat{\mu}_{\hat{\mathcal{G}}_1}\big(\cup_\alpha \hat{W}_{\alpha,1}: \mathbf{s}_1 = n\big) = d_0 q_n \tag{A.21}$$

and, similarly,

$$\hat{\mu}_{\hat{\mathcal{E}}_1}\big(\cup_\beta \hat{W}_{\beta,1,\mathbf{s}_1,*}: \mathbf{s}_1 = n\big) \geq d_0\, \hat{\mu}_{\hat{\mathcal{E}}_1}\big(\cup_\beta \hat{W}_{\beta,1}: \mathbf{s}_1 = n\big) = d_0 q_n \tag{A.22}$$

due to (A.18)–(A.20).

Next, for every $n \geq 1$ we again consider all the rectangles $\{\hat{W}_{\alpha,1}\}$ and $\{\hat{W}_{\beta,1}\}$ on which the function $\mathbf{s}_1$ takes value n. Their images at time n will contain H-components that fully cross the magnet $\mathfrak{S}$, and the relative measure of their intersections with $\mathfrak{S}$ will be $\geq d_0$ due to (A.21)–(A.22). At that time we apply our coupling procedure described in the first step and then link ('couple') their subsets of relative measure $d = d_0/2$, according to (A.12), so that

$$\hat{\mu}_{\hat{\mathcal{G}}_1}\big(\hat{\mathcal{G}}_2^{(n)}\big) = \hat{\mu}_{\hat{\mathcal{E}}_1}\big(\hat{\mathcal{E}}_2^{(n)}\big) = dq_n, \tag{A.23}$$

where

$$\hat{\mathcal{G}}_2^{(n)} := \big\{(x,t) \in \cup_\alpha \hat{W}_{\alpha,1,\mathbf{s}_1,*}: \mathbf{s}_1 = n \;\&\; \mathcal{F}^n(x,t) \text{ is coupled}\big\}$$

and

$$\hat{\mathcal{E}}_2^{(n)} := \big\{(y,t) \in \cup_\beta \hat{W}_{\beta,1,\mathbf{s}_1,*}: \mathbf{s}_1 = n \;\&\; \mathcal{F}^n(y,t) \text{ is coupled}\big\}.$$

Doing this for all $n \geq 1$ constitutes the second step of our construction. We denote by $\hat{\mathcal{G}}_2 = \cup_n \mathcal{F}^{-\mathbf{s}_0}\big(\hat{\mathcal{G}}_2^{(n)}\big)$ and $\hat{\mathcal{E}}_2 = \cup_n \mathcal{F}^{-\mathbf{s}_0}\big(\hat{\mathcal{E}}_2^{(n)}\big)$ the preimages of all the subsets 'coupled' during the second step of the construction. Note that $\hat{\mathcal{G}}_2 \subset \hat{\mathcal{G}}$ and $\hat{\mathcal{E}}_2 \subset \hat{\mathcal{E}}$. The coupling map Θ is thus extended to $\Theta: \hat{\mathcal{G}}_2 \to \hat{\mathcal{E}}_2$. We also define the coupling time function Υ on the set $\hat{\mathcal{G}}_2$ by

$$\Upsilon(x,t) = \mathbf{s}_0(x,t) + \mathbf{s}_1(\mathcal{F}^{\mathbf{s}_0}(x,t)).$$

It should be clear now how the construction of the coupling map proceeds as the above steps are repeated recursively.

Finally, we prove the clause B of Coupling Lemma 3.4 (this will also imply that the coupling map Θ is defined almost everywhere on the standard family

$\hat{\mathcal{G}}$). First we summarize the results of the previous constructions. For each $k \geq 1$, at the kth step we define the stopping time function $\mathbf{s}_{k-1}$ on the sets $\cup_\alpha \hat{W}_{\alpha,k-1}$ and $\cup_\beta \hat{W}_{\beta,k-1}$ of yet uncoupled points. Then we 'couple' some points of their images $\cup_\alpha \mathcal{F}^{\mathbf{s}_{k-1}}(\hat{W}_{\alpha,k-1})$ and $\cup_\beta \mathcal{F}^{\mathbf{s}_{k-1}}(\hat{W}_{\beta,k-1})$. Then we denote by $\hat{\mathcal{G}}_k$ and $\hat{\mathcal{E}}_k$ the preimages of just 'coupled' subsets, on the original families $\mathcal{G}$ and $\mathcal{E}$. Lastly we extend the coupling time function Υ to the set $\hat{\mathcal{E}}_k$ by

$$\Upsilon(z) = \mathbf{s}_0(z) + \mathbf{s}_1(\mathcal{F}^{\mathbf{s}_0}z) + \cdots + \mathbf{s}_{k-1}(\mathcal{F}^{\mathbf{s}_0+\cdots+\mathbf{s}_{k-2}}z),$$

where $z = (x, t) \in \hat{\mathcal{G}}_k$. Observe that the point $\mathcal{F}^{\Upsilon(z)}(z)$ and its partner $\mathcal{F}^{\Upsilon(z)}(\Theta(z))$ lie on the same stable H-manifold, which proves the claim A of Lemma 3.4 (assuming that Θ is indeed defined almost everywhere).

Next we rewrite the 'measure' relations (A.18)–(A.19) and (A.23) for the kth step. For brevity, we identify the set $\hat{\mathcal{G}}_k$ and $\hat{\mathcal{E}}_k$ with their images, i.e. we consider all our stopping time functions as defined on the original families $\hat{\mathcal{G}}$ and $\hat{\mathcal{E}}$. Then (A.18)–(A.19) generalize to the following 'conditional probability' formula

$$\hat{\mu}_{\mathcal{Y}}(\mathbf{s}_k = n/\mathbf{s}_{k-1} = n_{k-1}, \ldots, \mathbf{s}_1 = n_1, \mathbf{s}_0 = n_0) = q_n \tag{A.24}$$

where $\mathcal{Y} = \mathcal{G}$ or $\mathcal{E}$; and (A.23) generalize to another 'conditional probability' formula

$$\hat{\mu}_{\mathcal{Y}}\big(\hat{\mathcal{Y}}_k/\mathbf{s}_k = n_k, \ldots, \mathbf{s}_1 = n_1, \mathbf{s}_0 = n_0\big) = d, \tag{A.25}$$

where again $\mathcal{Y} = \mathcal{G}$ or $\mathcal{E}$.

The following argument is standard in the studies of random walks (but we do not assume here that the reader is familiar with it). Let

$$\bar{p}_n := \hat{\mu}_{\mathcal{G}}((x, t) \in \cup_\alpha \hat{W}_\alpha : \Upsilon(x, t) = n) \tag{A.26}$$

denote the fraction of points coupled exactly at time n (i.e., at the nth iteration of $\mathcal{F}$, rather than at the nth step of our construction). For example, $\bar{p}_i = 0$ for $i < n_0$ and $\bar{p}_{n_0} = d$. Then, due to (A.25), $p_n := \bar{p}_n/d$ will be the fraction of points *stopped* at time n, i.e.

$$p_n = \hat{\mu}_{\mathcal{G}}((x, t) \in \cup_\alpha \hat{W}_\alpha : \mathbf{s}_0 + \mathbf{s}_1 + \cdots + \mathbf{s}_k = n \text{ for some } k);$$

observe that p_n is not a probability distribution; in particular $p_{n_0} = 1$. Due to (A.24) and (A.25) we have the following 'convolution law':

$$p_{n+n_0} = (1-d)\Big(q_n + (1-d)\sum_{i=1}^{n-1} q_{n-i}\, p_{n_0+i}\Big) \qquad \forall n \geq 1. \tag{A.27}$$

Its verification is rather straightforward, and we leave it to the reader as an enjoyable exercise.

Now consider two complex analytic functions

$$P(z)=\sum_{n=1}^{\infty} p_{n_0+n}z^n \quad \text{and} \quad Q(z)=\sum_{n=1}^{\infty} q_n z^n.$$

Since $|p_n|\le 1$ and $|q_n|\le 1$, these functions are defined at least on the open unit disk $\{z\colon |z|<1\}$ in the complex plane $\mathbb{C}$. Moreover, (A.9) easily implies that $|Q(z)|\le 1$ for all $|z|\le 1$ and that $Q(z)$ is analytic in a slightly larger complex disk $\{z\colon |z|<\theta^{-1}\}$, here $\theta^{-1}>1$.

Equation (A.27) implies $P(z)=(1-d)\,Q(z)+(1-d)^2\,P(z)\,Q(z)$, hence

$$P(z)=\frac{(1-d)\,Q(z)}{1-(1-d)^2\,Q(z)}.$$

We see that $P(z)$ is analytic in some complex disk of radius greater than one, i.e. in $\{z\colon |z|<1+\delta\}$ for some $\delta>0$, hence $|p_n|\le \text{const}(1+\delta')^{-n}$ for any $\delta'<\delta$. A similar exponential bound then follows for $\bar{p}_n=d\,p_n$ introduced in (A.26). Coupling Lemma 3.4 is proved.

ACKNOWLEDGMENTS

This work is an extension of certain parts of the author's joint paper with D. Dolgopyat reported in [12]; many ideas behind these results actually belong to D. Dolgopyat. Also, R. Markarian has read main proofs and made useful remarks. The author is grateful to anonymous referees for many helpful comments. The author was partially supported by NSF grant DMS-0354775.

REFERENCES

1. R. Bowen, Equilibrium states and the ergodic theory of Anosov diffeomorphisms. *Lect. Notes Math.* **470** (Springer, Berlin, 1975).
2. X. Bressaud and C. Liverani, Anosov diffeomorphism and coupling. *Ergod. Th. Dynam. Syst.* **22**:129–152 (2002).
3. L. A. Bunimovich and Ya. G. Sinai, On a fundamental theorem in the theory of dispersing billiards. *Math. USSR Sbornik* **19**:407–423 (1973).
4. L. A. Bunimovich and Ya. G. Sinai, Markov partitions for dispersed billiards, *Commun. Math. Phys.* **73**:247–280 (1980).
5. L. A. Bunimovich and Ya. G. Sinai, Statistical properties of Lorentz gas with periodic configuration of scatterers. *Commun. Math. Phys.* **78**:479–497 (1981).
6. L. A. Bunimovich, Ya. G. Sinai and N. I. Chernov, Markov partitions for two-dimensional billiards. *Russ. Math. Surv.* **45**:105–152 (1990).
7. L. A. Bunimovich, Ya. G. Sinai and N. I. Chernov, Statistical properties of two-dimensional hyperbolic billiards, *Russ. Math. Surv.* **46**:47–106 (1991).
8. N. I. Chernov, Limit theorems and Markov approximations for chaotic dynamical systems. *Prob. Th. Rel. Fields* **101**: 321–362 (1995).
9. N. Chernov, Decay of correlations and dispersing billiards, *J. Statist. Phys.* **94**:513–556 (1999).

10. N. Chernov, Regularity of local manifolds in dispersing billiards. *Math. Phys. Electr. J.* **8**:1 (2006).
11. N. Chernov and C. Dettmann, The existence of Burnett coefficients in the periodic Lorentz gas. *Physica A* **279**:37–44 (2000).
12. N. Chernov and D. Dolgopyat, *Brownian Brownian Motion—I*, manuscript; available at http://www.math.uab.edu/chernov/papers/pubs.html
13. M. Denker, The central limit theorem for dynamical systems, *Dyn. Syst. Ergod. Th. Banach Center Publ.* **23** (PWN–Polish Sci. Publ., Warsaw, 1989).
14. C. Dettmann, The Burnett expansion of the periodic Lorentz gas. *Ergod. Th. Dynam. Syst.* **23**:481–491 (2003).
15. G. Gallavotti and D. S. Ornstein, Billiards and Bernoulli Schemes. *Commun. Math. Phys.* **38**:83–101 (1974).
16. I. A. Ibragimov and Y. V. Linnik, *Independent and stationary sequences of random variables* (Wolters-Noordhoff, Groningen, 1971).
17. V. P. Leonov, On the dispersion of time-dependent means of a stationary stochastic process. *Theory Probab. Its Appl.* **6**:87–93 (1961).
18. Ya. B. Pesin and Ya. G. Sinai, Gibbs measures for partially hyperbolic attractors. *Erg. Th. Dyn. Sys.* **2**:417–438 (1982).
19. W. Philipp and W. Stout, Almost sure invariance principles for partial sums of weakly dependent random variables. *Memoir. Amer. Math. Soc.* **161**: (1975).
20. D. Ruelle, A measure associated with Axiom-A attractors. *Amer. J. Math.* **98**:619–654 (1976).
21. D. Ruelle, *Thermodynamic Formalism* (Addison-Wesley, Reading, Mass, 1978).
22. Ya. G. Sinai, Markov partitions and C-diffeomorphisms. *Funct. Anal. Its Appl.* **2**:61–82 (1968).
23. Ya. G. Sinai, Dynamical systems with elastic reflections. Ergodic properties of dispersing billiards. *Russ. Math. Surv.* **25**:137–189 (1970).
24. Ya. G. Sinai, Gibbs measures in ergodic theory. *Russ. Math. Surveys* **27**:21–69 (1972).
25. L. Stojanov, An Estimate from above of the number of periodic orbits for semi-dispersed billiards. *Commun. Math. Phys.* **124**:217–227 (1989).
26. L.-S. Young, Statistical properties of dynamical systems with some hyperbolicity. *Annals of Math.* **147**:585–650 (1998).
27. L.-S. Young, Recurrence times and rates of mixing. *Israel J. of Mathematics* **110**:153–188 (1999).