Statistical properties of two-dimensional hyperbolic billiards

1991 Russ. Math. Surv. 46 47

(http://iopscience.iop.org/0036-0279/46/4/R02)

Uspekhi Mat. Nauk 46:4 (1991), 43–92 Russian Math. Surveys 46:4 (1991), 47–106

Statistical properties of two-dimensional hyperbolic billiards

L.A. Bunimovich, Ya.G. Sinai, and N.I. Chernov

Contents

§1. Introduction

The present paper is a continuation of the article [8] by the same authors, but can be read independently of it.

We will study the statistical properties of dynamical systems generated by a free motion with elastic reflections at the boundary. Such systems are called *billiards*.

Definition. Let Q be a bounded domain with piecewise smooth boundary in the Euclidean plane $\mathbb{R}^2$ or on the standard torus Tor^2. A *billiard* is a dynamical system generated by the motion of a point particle with constant unit velocity inside Q and with elastic reflections at the boundary ∂Q.

As usual, elastic reflection means reflection such that "the angle of incidence equals the angle of reflection".

Below we will consider *hyperbolic billiards* (see [34]), that is, billiards given inside a domain Q with boundary of a specific shape which preserves the hyperbolic nature of the motion (this means that the characteristic Lyapunov exponents are non-zero almost everywhere in the phase space).

More precisely, we will consider the class of two-dimensional billiards for which Markov partitions were constructed in [8]. These are billiards in domains whose boundary consists of finitely many C^3-smooth components of three shapes:

a) strictly concave (as seen from the inside of Q);

b) rectilinear segments;

c) convex (as seen from the outside) incomplete arcs of circles whose complements to complete circles do not intersect the other components of ∂Q.

The components of shape a) are called *scattering*, those of shape b) *neutral*, and those of shape c) *focussing* (the names are given according to their action on a pencil of parallel trajectories, see Fig. 1). Billiards with boundary consisting of components of shape a) only are said to be *scattering*, and consisting of components of shapes a) and b)—*semiscattering*. Among the concrete examples, the best known are the *periodic Lorentz gas* (a scattering billiard on the torus with cut out circles, Fig. 2a) and the *stadium* (a billiard in a domain bounded by two parallel segments and two arcs of circles, Fig. 2b).

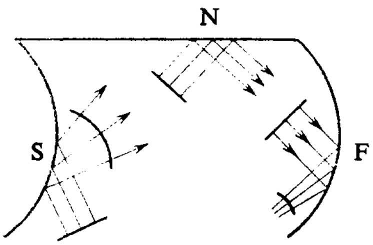

Fig. 1

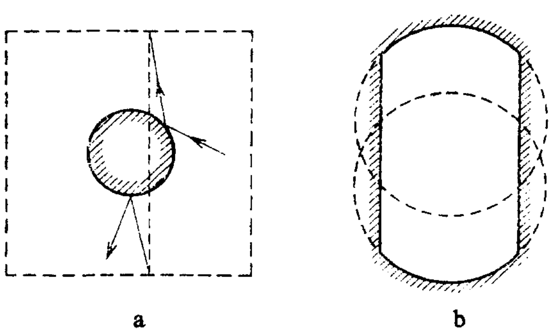

Fig. 2

A billiard system is Hamiltonian, and hence preserves the Liouville measure [10].

From a system in continuous time we can, in a standard manner, pass to the derived map T, corresponding to the map from a present reflection to the previous one (in the presence of focussing or neutral components of ∂Q it is constructed somewhat differently, see §2). The map T acts on the manifold of all states of the system "after reflection". This manifold is a two-dimensional surface M, which will be defined in a precise manner in §2. It will also be the phase space of T. The projection ν onto M of the Liouville measure will be an invariant measure for T [10]. In the presence of an invariant measure, every measurable function $F(x)$ on M gives rise to a stationary stochastic process in discrete time, $X_n = F(T^n x)$, $n \in \mathbb{Z}$. The statistical properties of this process (rate of decay of correlations, central limit theorem, and so on) play an important role in applications of the theory of billiards in physics. In this direction we will prove the following assertion (here $\langle\cdot\rangle$ denotes averaging with respect to the measure ν):

***Theorem* 1.1** (rate of decay of correlations). *Suppose we are given a two-dimensional hyperbolic billiard whose boundary satisfies certain additional conditions* A, B, C *of "general position" (see* §2*), while $F(x)$ is a function on M satisfying a Hölder condition in the natural coordinates r, φ (see also* §2*) with index bounded below*[(1)] *and $\langle F\rangle = 0$. Then the correlation of the stationary process $\{X_n\}$ decays subexponentially*:

$$(1.1) \qquad |\langle X_0 \cdot X_n\rangle| \leqslant C(F)\, e^{-a\sqrt{n}},$$

where $C(F) > 0$ and $a = a(Q) > 0$ is a constant.

If the index α in the Hölder condition for F is not bounded below, then Theorem 1.1 is also true, but with index a in (1.1) depending on α, and thus on F.

We put $S_n = X_1 + \ldots + X_n$.

***Theorem* 1.2** (central limit theorem). *Under the conditions of Theorem* 1.1 *the quantity*

$$(1.2) \qquad \sigma^2 = \sum_{n=-\infty}^{\infty} \langle X_0 \cdot X_n\rangle$$

is finite. If $\sigma \neq 0$, then

$$(1.3) \qquad \nu\left\{\frac{1}{\sigma\sqrt{n}} S_n < z\right\} \underset{n\to\infty}{\longrightarrow} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-u^2/2}\, du.$$

Remark 1.3. A more correct name for Theorem 1.1 would have been the theorem on the decay of *autocorrelations*. By *correlation* proper we often mean the quantity $\langle F(x)\cdot G(T^n x)\rangle$, where F and G are distinct functions on M satisfying the same conditions as F does in Theorem 1.1. Here we can prove the estimate

$$(1.4) \qquad |\langle F(x)\cdot G(T^n x)\rangle| \leqslant C(F, G)\, e^{-a\sqrt{n}}.$$

Remark 1.4 [9]. The quantity σ^2 in (1.2) is equal to zero if and only if the function $F(x)$ is homologous to zero, that is, $F(x) = G(Tx) - G(x)$, where G is a function on M that is square integrable with respect to the measure ν.

We will also give two specific consequences of Theorem 1.2, which have a clear physical interpretation. Both consequences relate to the more narrow class of systems of *scattering billiards with finite horizon* (that is, when the length of a free path between reflections is uniformly bounded above; this condition can be violated only in billiards on the torus).

Let $N_t(x)$ be the number of reflections experienced by the trajectory of a point $x \in M$ from time 0 till time t (this is the continuous time of the phase flow).

---

(1)This means that $|F(x) - F(y)| \leqslant C(F)|x-y|^{\alpha}$ when $\alpha \geqslant \alpha_0$, for some $\alpha_0(Q) > 0$.

***Theorem* 1.5.** *There are numbers $a_1 = a_1(Q) > 0$ and $b_1 = b_1(Q) > 0$ such that*

$$\nu\left\{\frac{N_t(x) - a_1 t}{\sqrt{b_1 t}} \leqslant z\right\} \underset{t\to\infty}{\longrightarrow} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-u^2/2}\, du. \tag{1.5}$$

The second consequence relates to the appearance of diffusion in the periodic Lorentz gas. Let Q be a domain on the standard torus with coordinates $0 \leqslant q_1 < 1$ and $0 \leqslant q_2 < 1$. We denote by Q_∞ the corresponding domain on the universal enveloping torus: $Q_\infty = \{(q_1, q_2) \in \mathbb{R}^2 : (q_1 - n, q_2 - m) \in Q$ for certain integers $m, n\}$. Then the lift of any trajectory of a billiard onto the universal envelope will be a trajectory of a billiard in Q_∞. If the initial set of trajectories (for $t = 0$) is concentrated in the initial domain $Q = Q_\infty \cap \{0 \leqslant q_1, q_2 < 1\}$ and is distributed in accordance with the Liouville measure, then as $t \to \infty$ it will "unravel" (diffuse) into Q_∞. We denote by $q_1^{(t)}(x)$ and $q_2^{(t)}(x)$ the coordinates (in Q_∞) of the wandering point at the moment of time t whose initial position (for $t = 0$) was at $x \in M$.

***Theorem* 1.6.** *There is a two-dimensional Gaussian distribution with density $g(q_1, q_2)$ (depending on the domain Q) such that*

$$\nu\left\{\left(\frac{q_1^{(t)}(x)}{\sqrt{t}}, \frac{q_2^{(t)}(x)}{\sqrt{t}}\right) \in A\right\} \underset{t\to\infty}{\longrightarrow} \int_A g(q_1, q_2)\, dq_1\, dq_2 \tag{1.6}$$

for each bounded open set $A \subset \mathbb{R}^2$ whose boundary has measure 0. This Gaussian distribution has zero mean and non-singular covariance matrix.

We will briefly clarify the idea behind the proofs of Theorems 1.1 and 1.2. Statistical properties of hyperbolic dynamical systems were first studied in the case of geodesic flows on manifolds of negative curvature [1], and subsequently for the axiomatically defined Anosov Y-systems [12] and the more general A-systems of Smale [35], [19], [20]. The general scheme of investigation consists in constructing a Markov partition, with subsequent reduction of the system to its symbolic representation as a topological Markov chain (TMC) with finite alphabet. For a TMC having the additional property of topological mixing we can define the large class of invariant Gibbs measures, which are the equilibrium states corresponding to "good" (usually, Hölder) functions on the phase space. The statistical properties of Gibbs measures are sufficiently well investigated in the work of Bowen, Ruelle, and Sinai; see the surveys [4], [33], [14]. The corresponding theory, which has received the appellation "thermodynamical formalism", has deep connections with the one-dimensional statistical mechanics of lattice systems. In particular, in the above-cited works the exponential nature of decay of correlations and the central limit theorem were proved. We also note the recent paper [28], in which these results were derived anew by using purely probabilistic techniques (by means of reduction to general Markov chains), and, moreover, in which a local limit theorem and a renewal theorem were proved.

The billiard systems considered in the present paper are hyperbolic, and, moreover, to a sufficiently high degree: their hyperbolicity is close to uniform, and since the phase space is two-dimensional, they are complete (for details see §2). However, billiard systems differ in two essential aspects from Y- and A-diffeomorphisms.

First, the billiard map T is discontinuous. Its discontinuities are generated by tangent (sliding) maps and the tangential coincidence of trajectories at a break point of ∂Q. Therefore the elements of the Markov partition have a rather complicated shape: they are nowhere dense and form totally disconnected sets of Cantor type. More important, however, is that they are countable in number, hence we obtain a symbolic system with infinite alphabet (see [8] for a more detailed explanation).

The second difference is the fact that in billiards there is a natural invariant measure ν, induced by the Liouville measure. From the point of view of applications this measure is very interesting. It induces in the symbolic system an invariant measure ν', which by no means needs to be Gibbsian. In passing we note that the construction of a Gibbs measure for a TMC with countable alphabet is a problem which has not yet been completely solved.

These two circumstances impel us to work out a new approach to the investigation of statistical properties of billiard systems. The first such investigation was done in [22], [23]. In these papers the measure ν' in the TMC was approximated by a measure ν'' with finite memory, that is, by an "almost" Markov measure. The resulting probability chain can be easily reduced to a Markov chain (but with countably many states). For proving statistical properties of the latter, "simple" mixing is insufficient, since this may happen too slowly, and an additional regularity condition is necessary. One such condition, an analogue of Doeblin's condition, was proved in [22] (see the correction in [24]), and as a result in [23] the authors obtained a subexponential estimate for the decay of correlations:

$$|\langle X_0 \cdot X_n \rangle| \leqslant C(F)\, e^{-n^{\gamma}} \tag{1.7}$$

(here $\gamma \in (0, 1)$ is independent of F), and proved a central limit theorem and a number of other statistical properties for the periodic Lorentz gas under the condition of finite horizon (with an upper bound on the length of a free path). On the basis of these results, in [30] a "quasilocal" limit theorem was obtained, taking an intermediate position between central and local.

In the present paper we have substantially simplified this approach and have obtained stronger results. First we get rid of the construction of a symbolic system with countable alphabet, which allows one to code almost all points of the phase space M. In fact, when approximating the measure ν' by the measure ν'' in the TMC some loss of exactness occurs, and this can be allowed for from the very beginning, by constructing the symbolic system more "roughly", but with finite alphabet. For this we introduce a finite family of subsets of M which have the Markov property during a definite *finite* interval of time (that is, for the action of T^i with $0 \leqslant i \leqslant N$, where N is finite). This family of subsets *does not cover* all of the phase space, hence we call it a *Markov lattice*. It generates a symbolic (non-Markovian!) system with finite alphabet whose invariant measure can be sufficiently well approximated by a Markov measure ν'' (again, on the same time interval $[0, N]$). This measure has good statistical properties, not "everywhere" but precisely on the finite time interval $[0, N]$. In this setting we can prove, as in [9], the additional regularity condition. We will establish an analogue of the condition of strong mixing in the sense of Ibragimov [9] (this is stronger than the Doeblin condition), and derive Theorems 1.1 and 1.2 from it. Note that our estimate (1.1) is better than (1.7), since one can actually estimate γ from below, which gives $\gamma \geqslant 1/2$. Moreover, our assumptions A, B, and C on the boundary ∂Q are essentially weaker than those given in [23] (for example, here we consider the Lorentz gas with infinite horizon, and also the stadium).

The question whether the subexponential estimate of the form (1.1) cannot be improved upon for billiards remains open. However, numerical calculations [18], [26] indicate that the actual nature of decay of correlations is *subexponential.*

Finally we note that in our proofs of Theorem 1.1 and 1.2 the Markov partition is superfluous: our lattice is constructed using the *pre-Markov* partition introduced in [18], [22] as an intermediate stage in the construction of a Markov partition. In essence, this also simplifies the proofs of Theorems 1.1 and 1.2 in comparison with [22], [23].

On the other hand, in billiards the Markov partition is necessary in the solution of a number of other problems. It allows us to prove certain asymptotic estimates for the number of periodic trajectories (see [8], [17]). It may be expected that it also allows us to develop a thermodynamical formalism in the spirit of Bowen–Ruelle–Sinai for hyperbolic systems with singularities (in particular, for hyperbolic billiards).

The structure of the present article is as follows. In §2 we briefly describe the general properties of hyperbolic billiards. In §3 we develop the necessary technique for constructing a Markov lattice, based on the notion of a *homogeneous* stable or unstable manifold, which allows us to overcome the influence of the singularities of the billiard map T. In §4 we construct the Markov lattice, and state its definition and basic properties in §4.1. The last three sections are devoted to the proof of Theorems 1.1 (§5), 1.2 (§6), and 1.5–1.7 (§7). The Appendices contain the proofs of certain technical theorems in §3.

The notion of a Markov lattice was proposed by N.I. Chernov. The idea of passing to homogeneous manifolds, and the related additional partitions (see §3), was coined by Ya.G. Sinai. The final editing was done by all the authors.

§2. Billiards: necessary information

In this section we state the necessary definitions and properties of hyperbolic billiards. For a more detailed account see [8].

2.1. Scattering billiards: the phase space.
As in [8] we direct our main attention to scattering billiards, and state for the remaining billiards only the necessary changes and additions in all constructions.

A billiard domain Q is always assumed to be a bounded closed connected domain in the plane or on the two-dimensional standard torus. The boundary ∂Q must be piecewise smooth and consist of finitely many smooth (of class at least C^3) non-self-intersecting curves Γ_i, $1 \leqslant i \leqslant d$, which are either closed or have end-points in common. The *regular part of the boundary* is denoted by

$$\widetilde{\partial Q} = \partial Q \setminus \bigcup_{i \neq j} (\Gamma_i \cap \Gamma_j),$$

while a point $q \in \partial Q \setminus \widetilde{\partial Q}$ is called a *break point of it*. At regular points $q \in \widetilde{\partial Q}$ the unit interior normal vector $n(q)$ and curvature $\varkappa(q)$ with respect to this vector are defined. In scattering billiards $\varkappa(q)$ is everywhere positive.

A billiard system specifies a piecewise smooth flow $\{S^t\}$ on the phase space $\mathfrak{M}$, which can be represented in the form $\mathfrak{M} = Q \times S^1 = \{x = (q, \upsilon) : q \in Q, \|\upsilon\| = 1\}$. The flow $\{S^t\}$ preserves the Liouville measure $d\mu = c_\mu dq d\upsilon$, where dq and $d\upsilon$ are the Lebesgue measures on Q and S^1, respectively; c_μ is a normalizing factor.

We introduce the "restricted phase space" $M = \{x = (q, \upsilon) : q \in \widetilde{\partial Q}, (\upsilon, n(q)) > 0\}$. We let $\overline{M}$ be the closure of M in $\mathfrak{M}$. The boundary $\partial M = \overline{M} \setminus M$ consists of two parts: $\partial M = S_0 \cup V_0$, where $S_0 = \{(q, \upsilon) : q \in \widetilde{\partial Q}, (\upsilon, n(q)) = 0\}$ ("tangential reflections") and $V_0 = \{(q, \upsilon) : q \in \partial Q \setminus \widetilde{\partial Q}\}$ ("billiard corners"). The flow $\{S^t\}$ induces a derived map from $\overline{M}$ into itself, denoted by T.

We introduce in $\overline{M}$ natural coordinates: r, the arc length parameter on the curve ∂Q, and φ, the angle between the vectors υ and $n(q)$ ($|\varphi| \leqslant \pi/2$). In these coordinates $\overline{M}$ is the union of rectangles and cylinders. For a point $x \in \overline{M}$, we denote its coordinates by $r(x)$ and $\varphi(x)$. The map T preserves the measure $d\nu = c_\nu \cos\varphi \, dr d\varphi$ (c_ν is a normalizing factor). We denote by $\tau_+(x) = \tau(x)$ and $\tau_-(x)$ the first positive and negative moments when the trajectory of x hits the boundary ∂Q, that is, $T^{\pm 1}x = S^{\tau_\pm(x)+0}x$.

The maps T and T^{-1} are piecewise smooth: T has discontinuities on the set $T^{-1}R_0$, and T^{-1} on TR_0. We put $R_i = T^i R_0$ and $R_{m,n} = \bigcup_m^n R_i$ for $-\infty \leqslant m < n \leqslant \infty$. Then the set of singular points for $T^{\pm n}$, $n \geqslant 1$, coincides with $R_{-n,0}$ ($R_{0,n}$). The set $R_{-\infty,\infty}$ consists of countably many smooth (C^1) curves, called *discontinuity curves* in the sequel.

On a scattering billiard we impose the following two conditions of "general position".

Condition A. *All interior angles of the domain Q at the break points of its boundary are distinct from zero.*

Condition B. *For any $m \geqslant 1$, the number of discontinuity curves in $R_{-m,m}$ passing through or ending at some point $x \in \overline{M}$ does not exceed $K_0 m$, where K_0 is a constant for the domain Q.*

Both conditions were used in [8] in an essential manner in the construction of Markov partitions. Note that condition B is always fulfilled if Q is a domain with smooth boundary on the torus [8]. In [22], [23] condition B is formulated differently, and more rigidly: it is not allowed that more than three discontinuity curves from $R_{-\infty,\infty}$ pass through a point of $\overline{M}$. Numerical computations [32] indicate that when condition A is violated, the rate of decay of correlations becomes slower (power-like), that is, Theorem 1.1 is not valid.

Remark 2.1. Suppose the billiard domain Q satisfies condition A. Then there are constants $m_0 = m_0(Q)$ and $\tau_0 = \tau_0(Q) > 0$ such that for any point $x \in \overline{M}$ there is among its images $T^i x$ for $1 \leqslant i \leqslant m_0$ at least one for which $\tau(T^i x) > \tau_0$.

In other words, the trajectories of a billiard cannot undergo arbitrarily many reflections while remaining in a small neighbourhood of one of the break points of the boundary ∂Q.

Moreover, we note that since the boundary components are smooth, the curvature is bounded above and below: $0 < \varkappa_{\min} \leqslant \varkappa(x) \leqslant \varkappa_{\max} < \infty$.

2.2. Scattering billiards: the hyperbolic structure.
A smooth (C^1) curve $\gamma \subset \overline{M}$ is said to be *increasing* (*decreasing*) if it is given by an equation $\varphi = \varphi(r)$ and $d\varphi/dr > 0$ ($d\varphi/dr < 0$). The property of being increasing (decreasing) is preserved under the action of T (T^{-1}).
A curve γ is said to be *m-increasing* (*m-decreasing*) for $m \geqslant 1$ if $T^{-m}\gamma$ ($T^m\gamma$) is an increasing (decreasing) smooth curve. All discontinuity curves in $R_{1,\infty}$ ($R_{-\infty,-1}$) are increasing (decreasing).

Let γ be an increasing or decreasing curve, given by the equation $\varphi = \varphi(r)$. We denote by $l(\gamma)$ its Euclidean length in the coordinates (r, φ). The quantity

$$p(\gamma) = \int_\gamma \cos\varphi \, dr \tag{2.1}$$

is called the *p-length of the curve* γ, and for us it will be far more important than the l-length. In the sequel, unless otherwise stated, the length of a curve will mean this p-length.

All increasing (decreasing) curves expand under the action of T (T^{-1}) (in the sense of p-length). The local coefficient of expansion of increasing (decreasing) curves under the action of T^n for $n \geqslant 1$ ($n \leqslant -1$) always grows exponentially:

$$p\,(T^n\gamma)/p\,(\gamma) \geqslant \Lambda_0^{[|n|/m_0]} \tag{2.2}$$

for some $\Lambda_0(Q) > 1$. By the same token, the hyperbolicity of the transformation T is close to *uniform*, and since the phase space $\overline{M}$ is two-dimensional, T is *complete* (for the definition, see [16]). However, T is not uniformly hyperbolic, since angles between m-increasing and m-decreasing curves are not bounded away from zero for any $m \geqslant 1$. More precisely, for any 1-increasing or 1-decreasing curve, $|d\varphi/dr| \geqslant \text{const} > 0$, but the upper bound is very weak: $|d\varphi/dr| \leqslant \text{const}(d(r))^{-1/2}$, where $d(r)$ denotes the distance from the point $r \in \Gamma_i \subset \partial Q$ to the nearest end-point of the curve Γ_i. This means that in a neighbourhood of the set V_0 the stable and unstable directions can be simultaneously almost parallel to the tangent to V_0. More precisely, this behaviour is observed in a neighbourhood of *finitely many points* only, lying in $V_0 \setminus S_0$, but it complicates our exposition in Appendix 1 (§A1.3).

Note that the l- and p-lengths of m_0-increasing and m_0-decreasing curves γ are related by

$$\text{const}_1\,(Q)\,p\,(\gamma) \leqslant l\,(\gamma) \leqslant \text{const}_2\,(Q)\,\sqrt{p\,(\gamma)}. \tag{2.3}$$

For almost every point $x \in \overline{M}$ there are *locally stable* and *unstable manifolds* (LSM and LUM, for short) passing through x. We denote the maximal smooth segment (including end-points) of this LUM (LSM) by $\gamma^u(x)$ ($\gamma^s(x)$). The lengths of these segments are bounded, because T is discontinuous, and, moreover, arbitrarily short LUM and LSM are everywhere dense in M.

The tangent directions to the LUM $\gamma^u(x)$ and to the LSM $\gamma^s(x)$ at x form angles with the r-axis, which we denote by $\psi^u(x)$ and $\psi^s(x)$, respectively. They are given by the relations

$$\begin{cases} \tan\psi^u(x) = B^u(x)\cos\varphi(x) + \varkappa(x), \\ \tan\psi^s(x) = -B^s(x)\cos\varphi(x) + \varkappa(x). \end{cases} \tag{2.4}$$

Here

$$B^s(x) = \cfrac{1}{\tau(x) + \cfrac{1}{R(Tx) + \cfrac{1}{\tau(Tx) + \cfrac{1}{R(T^2x) + \dots}}}} \tag{2.5}$$

is a continued fraction in which $R(x) = 2\varkappa(x)/\cos\varphi(x)$. The quantity $B^u(x)$ is similarly defined, using the semitrajectory $T^n x$ for $n \leqslant 0$:

$$B^u(x) = R(x) + \cfrac{1}{\tau(T^{-1}x) + \cfrac{1}{R(T^{-1}x) + \cfrac{1}{\tau(T^{-2}x) + \dots}}}$$

The value of $B^u(x)$ ($B^s(x)$) equals the curvature of the pencil of trajectories forming the LUM $\gamma^u(x)$ (LSM $\gamma^s(x)$) at the moment *after* reflection at x. We also denote by $B^u_-(x) = B^u(x) - R(x)$ and $B^s_-(x) = B^s(x) + R(x)$ the curvatures of these pencils *before* reflection at x. Incidentally, the p-length of the LUM $\gamma^u(x)$ (LSM $\gamma^s(x)$) defines the length of an orthogonal cross-section of the corresponding pencil of trajectories *at the moment of reflection.* We denote by $\lambda^u(x) = 1 + \tau(x)B^u(x)$ ($\lambda^s(x) = 1 + \tau(x)B^s_-(Tx)$) the coefficient of expansion of the LUM $\gamma^u(x)$ (coefficient of contraction of the LSM $\gamma^s(x)$) under the action of T (in the sense of p-length).

For $x, y \in \overline{M}$ we put $[x, y] = \gamma^u(x) \cap \gamma^s(x)$ (this point does not always exist, since the LUM and LSM may be arbitrarily short). For $A, B \subset \overline{M}$ we put $[A, B] = \{[x, y] : x \in A, y \in B\}$. For $A \subset \overline{M}$ we put $\gamma^{u,s}_A(x) = A \cap \gamma^{u,s}(x)$.

Two subsets $A \subset \gamma^u(x_0)$ and $B \subset \gamma^u(y_0)$ are said to be *canonically isomorphic* if for any $x \in A$ the point $[x, y_0] \in B$, and conversely. In this case the map $A \to B$ mapping x to the point $[x, y_0]$ is called the *canonical isomorphism.* If the p-length is considered as a *measure* on the segments $\gamma^u(x_0)$ and $\gamma^u(y_0)$, then the canonical isomorphism is absolutely continuous [13], [29], and has a Jacobian at every density point x of A, which we denote by $J^u(x)$. For any $A \subset \overline{M}$ and LUM $\gamma^u(x_0)$, we call $[A, x_0]$ the *canonical projection from the set A onto the curve $\gamma^u(x_0)$.*

2.3. Billiards with infinite horizon.

If Q is a domain on the torus, then the length $\tau(x)$ of a free path may be unbounded. In this case the functions $\tau_\pm(x)$ have singularities, which we will describe in more detail. The function $\tau_+(x)$ ($\tau_-(x)$) can be unbounded only in a neighbourhood of finitely many points lying in R_0, called *u-singularities* (*s-singularities*). They are also called *singular points of infinite horizon type* (we will encounter another type of singular points in §2.5). These singular points can be of three subtypes: S, V, and SV, depending on whether they belong to $S_0 \setminus V_0$, $V_0 \setminus S_0$, or $S_0 \cap V_0$ (for more details see [8], §4). Each subtype can be characterized by the structure of the discontinuity curves in R_{-1} (in R_1 for s-singularities), which accumulate in a neighbourhood of the singular point. Note that if the boundary ∂Q is smooth, then the u- and s-singular points coincide, and have one type: S.

The discontinuity curves partition the neighbourhood of a singular point into countable many domains (cells). When approaching the singular point, the cells become smaller, the functions $\tau_\pm(x)$ grow in them, and the coefficient of expansion of the LUM (coefficient of contraction of the LSM) increases to infinity by one iteration of $T^{\pm 1}$. If we number the cells (and the bounding discontinuity curves) in the natural order (that is, in the order of approach to the singular point [8]), then in the n-th cell the functions $\tau_\pm(x) \approx \text{const}\cdot n$, while the coefficient of expansion of the LUM (coefficient of contraction of the LSM) is approximately equal to $\text{const}\cdot n^d$, where $d = 3/2$ for singular points of types S and SV and $d = 1$ for singular points of type V.

Remark 2.2. In neighbourhoods of singular points of type V, any LUM intersects only finitely many curves in R_{-1} with indices N_1 to N_2 such that $N_2 \leqslant \text{const} \cdot N_1$. We also note that already the images and inverse images of cells from a neighbourhood of the singular points of type V do not lie in neighbourhoods of these or other singular points (hence in [8] they were called "wandering cells"). More precisely, the cells with indices $\geqslant n$ require at least $N(n)$ iterations of $T^{\pm 1}$ for their images to intersect other cells with indices $\geqslant n$ around singular points (here $N(n) \to \infty$ as $n \to \infty$).

2.4. Semiscattering billiards.
In semiscattering billiards the reflections in neutral boundary components are at most a "disturbing factor", since they do not lead to contraction and expansion. To exclude the influence of this factor we go over to a derived map on a "smaller" (than in §2.1) restricted phase space. We denote by $\partial^+ Q$ the union of all scattering components of the boundary ∂Q, and consider the space $M = \{(q, v) : q \in \partial^+ Q, (v, n(q)) > 0\}$. On the closure $\overline{M}$ we construct the derived automorphism T. For a point $x \in \overline{M}$ we denote by $k(x)$ the exponent of the first reflection of the trajectory of x in $\partial^+ Q$. We impose the additional condition:

***Condition* C.** *The function* $k(x)$ *is uniformly bounded*: $k(x) \leqslant \text{const} < \infty$.

Under this condition the discontinuity curves R_1 and R_{-1} can accumulate only in neighbourhoods of singular points of infinite horizon type.

In certain cases we can discard the rather stringent condition C, and by some reflections in Q relative to the neutral components of ∂Q reduce the system to a scattering billiard on the torus (see [8], §6). These cases are also very convenient for us, since the map T will have the same properties as in the corresponding scattering billiard.

However, if in a domain Q in general position condition C is violated, then the description of the structure of the discontinuity curves of T in neighbourhoods of singular points (in which the function $k(x)$ is unbounded) is up to now an unsolved problem.

2.5. Billiards with focussing boundary components.
On the focussing components of the boundary ∂Q we impose the following conditions:

***Condition* F1.** *Every focussing boundary component is an incomplete arc of a circle whose complement to a full circle does not intersect other boundary components.*

***Condition* F2.** *If a focussing component intersects a scattering one, then at their point of intersection they form an interior angle larger than* π.

***Condition* F3.** *Every focussing component is not larger than a semicircle (of angular measure* $\leqslant \pi$*).*

In [8], in the construction of a Markov partition we have stated clearly only condition F1, but implicitly assumed conditions F2, F3 also (since when these conditions are violated the discontinuity curves of the derived map T become different from those described in [8]).

The hyperbolic and ergodic properties of billiards with focussing components satisfying condition F1 were studied in [6], [21], and described in [7], [8]. When hitting the focussing part of the boundary ∂Q the images of increasing curves become decreasing in the coordinates r, φ, and conversely. Pencils of trajectories which are images of increasing curves converge (focus) after reflection in focussing components of the boundary, and before the next reflection they pass through a conjugate point (defocus) and approach this reflection in scattered form already. It is important that the conjugate point always lies on the first half of the path between adjacent reflections, hence the dimensions (p-lengths) of the images of increasing curves grow monotonically under the action of T. This phenomenon is called the *defocussing condition*.

A series of successive reflections in one focussing boundary component can be arbitrarily long, but it does not lead to exponential expansion and contraction (on this ground, in particular, it follows that billiards in the circle are not hyperbolic). These series become a "disturbing factor" similar to reflections in neutral boundary components, hence we have to "suppress" them too. More precisely, we let M be the set of points $\{(q, v): q \in \widetilde{\partial Q}, (v, n(q)) > 0\}$ such that the point q lies either on a scattering component or on a focussing component, and in the latter case the next reflection (at $S^{\tau(x)+0}x$) must happen in *another* boundary component. We construct the derived automorphism T on the closure $\overline{M}$. If the boundary contains neutral components, we also impose condition C.

With the coordinates r, φ the space $\overline{M}_-$ is the union of rectangles and cylinders, corresponding to scattering boundary components (we denote this part of M by M_+), and parallelograms, corresponding to focussing components (we denote this part of M by M_-). The parallelograms in M_- have the shape depicted in Fig. 3. At points A and B infinitely many discontinuity curves accumulate (they are described in more detail in [8]). The limit points A and B are given by the tangential directions to the corresponding focussing component Γ at its end-points. Figuratively speaking, the trajectories of these points are not reflected in the arc Γ but "slide" along it, hence we call A and B *singular points of sliding type*.

A neighbourhood of the singularities just described is partitioned by the curves in R_1 into countably many cells. In the n-th cell those trajectories are collected that have undergone n successive reflections in the given focussing component. Any LUM γ_0 with end-points x and y and lying in the n-th cell has derivative $d\varphi/dr \approx \text{const}\cdot n$ (of course, const < 0); for it $\cos\varphi(x) \approx \cos\varphi(y) \approx \text{const}/n$ and $|\varphi(x)-\varphi(y)| \leqslant \text{const}/n^2$. This easily implies the estimates

$$|r(x) - r(y)| \leqslant \text{const}\sqrt{p(\gamma_0)}, \tag{2.6}$$

$$\frac{|\varphi(x) - \varphi(y)|}{\cos\varphi(x)} \leqslant \text{const}\,(p(\gamma_0))^{1/3}, \tag{2.7}$$

which we will need in Appendix 1. It can also be easily proved that $p(\gamma_0)/p(T^{-1}\gamma_0) \approx \text{const}\cdot n$ and $p(T\gamma_0)/p(\gamma_0) \approx \text{const}\cdot n$, hence the coefficient of expansion of the LUM after *two* iterations of the map T (from $T^{-1}\gamma_0$ to $T\gamma_0$) is at least $\text{const}\cdot n^2$.

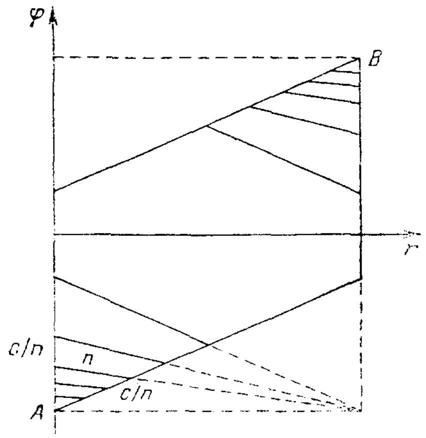

Fig. 3

2.6. The "stadium".

In §2.5, as in §2.4, condition C is very stringent. Here we consider a domain not satisfying it, the so-called *stadium*. It is bounded by two parallel segments and two similar circular arcs (see Fig. 2b). The complement of each arc to a full circle is not to intersect or touch the other arc. Billiards in a stadium have applications of their own (see [7]). The hyperbolicity and ergodicity of this billiard were proved in [21]. If the arcs bounding the stadium are smaller than a semicircle, then they do not satisfy condition F1, and by identifying the lateral walls (the segments) we may pass to a billiard on the torus, bounded by two arcs only, for which F1 does hold.

The restricted phase space M consists of two identical parallelograms (Fig. 4, see p. 60). In neighbourhoods of the points C and D there accumulate also infinitely many discontinuity curves (from R_{-1}), whose structure has been described in [8]. They are generated by a free path of unbounded length (under transition of the trajectories from one arc to the other), hence we will call C and D, as in §2.3, *singular points of infinite horizon type*. If the boundary arcs of the stadium are strictly smaller than a semicircle, then the properties of these singular points are completely similar to the properties of singular points of type V (see Remark 2.2 about "wandering cells").

If the stadium is bounded by semicircles, the cells in neighbourhoods of C and D are already non-wandering: the discontinuity curves from R_1 and R_{-1} have their location there (Fig. 5). In this case a simple computation gives the following estimate:

***Lemma* 2.3.** *The coefficient of expansion under the action of T of an arbitrary* LUM *lying in the n-th cell is at least* $4n$. *Any* LUM *in a neighbourhood of the point C or D intersects only finitely many discontinuity curves from R_{-1} with indices between N_1 and N_2, where* $N_2 \leqslant 9N_1 + \text{const}$.

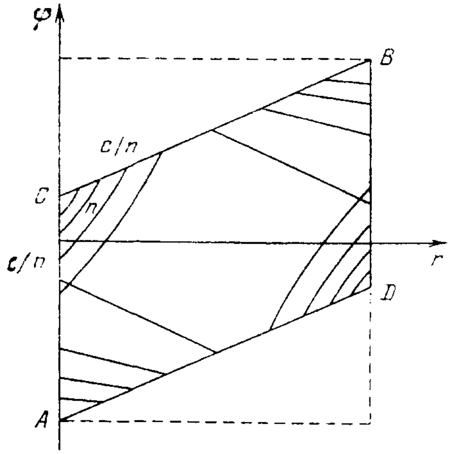

Fig. 4

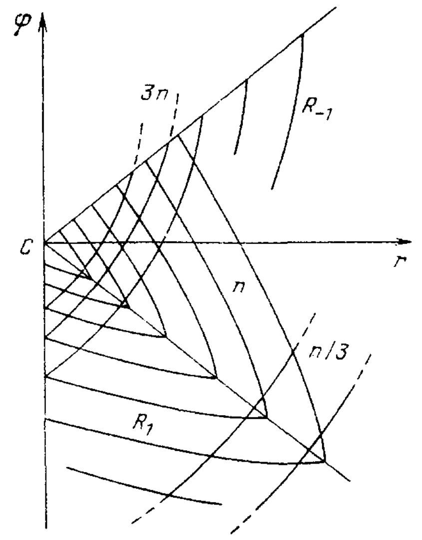

Fig. 5

In conclusion we make another remark to §2.4–§2.6.

Remark 2.4. For billiards with neutral and focussing boundary components, as in [8] we have constructed the derived map T on a very "small" restricted phase space M. In [8] we could then define a Markov partition also on the larger (and natural) space $M_1 = \{(q, v) : q \in \partial\widetilde{Q}, (v, n(q)) > 0\}$. In this paper we study much more refined statistical properties of dynamical systems, which may vary strongly under the transition from T to the derived map T_1 on M_1. For example, the "counter" of the number of collisions ($T^n x = T_1^N x$ for $N \geqslant n$, where N can be *much* larger than n) essentially changes, hence the rate of decrease of correlations may change substantially. In particular, as shown by numerical calculations in [36], for the stadium it becomes power-like ($\sim \text{const} \cdot N^{-a}$ for some $a > 0$).

§3. The Markov lattice: local construction

In this section we prepare the necessary material for the construction of the Markov lattice, which will be given in §4. We will also introduce a number of new notions and prove a number of results on the behaviour of the map T. When possible we will give our results an intuitive appearance, transferring the technical exposition to the Appendices.

3.1. Parallelograms and their measures.
By tradition, the elements of Markov partitions in hyperbolic systems are called *parallelograms*. In the case of billiards a representation by a parallelogram is a set $U \subset \overline{M}$ such that $\nu(U) > 0$ and for every pair $x, y \in U$ the point $[x, y]$ is defined and belongs to U. In other words, it is the set obtained by intersecting the family of LUM $\{\gamma^u\}$ by the family of LSM $\{\gamma^s\}$ such that *each* LUM γ^u intersects *each* LSM γ^s. In this sense the parallelogram U has the structure of a direct product. Hence U can also be represented in the form

$$U = [\gamma_U^u(x_0), \gamma_U^s(x_0)] \tag{3.1}$$

for an arbitrary point $x_0 \in U$. If x_0 is fixed, the LUM $\gamma^u(x_0)$ and the LSM $\gamma^s(x_0)$ will be called *coordinate axes in U.*

Because of the presence of discontinuities in billiard systems, arbitrarily short LUM and LSM are everywhere dense in M (for more detail see [8]). Hence the parallelograms form nowhere dense sets of Cantor type, having positive measure. The limit (in the metric of C^0) of the sequence of LUM or LSM can only be an LUM or LSM (see [8], Lemma 2.11), hence the closure (in $\overline{M}$) of any parallelogram is also a parallelogram.

We will need convenient formulae for the measures of parallelograms. Let U be a parallelogram, U_0 some "ambient" parallelogram, that is, $U_0 \supset U$, and let a point x_0 be chosen in U_0, fixing the coordinate axes in U_0. We denote the canonical projections of U onto the axes $\gamma^u(x_0)$ and $\gamma^s(x_0)$ by Γ_U^u and Γ_U^s, respectively. Then, by analogy with (3.1),

$$U = [\Gamma_U^u, \Gamma_U^s]. \tag{3.2}$$

We will partition the curves $\gamma^u(x_0)$ and $\gamma^s(x_0)$ into fine subsegments $\Delta_i^u \subset \gamma^u(x_0)$ and $\Delta_j^s \subset \gamma^s(x_0)$. Their "direct products" $[\Delta_i^u, \Delta_j^s]$ give a partition of U into parallelograms $\Delta_{ij} = [\Delta_i^u \cap \Gamma_U^u, \Delta_j^s \cap \Gamma_U^s]$.

The ν-measures of sufficiently fine parallelograms Δ are approximately equal to $\nu(\Delta) \approx c_\nu m(\Delta) \cos\varphi(x)$, where $m(\cdot)$ is Lebesgue measure in the coordinates r, φ and $x \in \Delta$ is some point. If the point $x \in \Delta$ is fixed, we may write $m(\Delta) \approx l(\gamma_\Delta^u(x)) \cdot l(\gamma_\Delta^s(x)) \cdot \sin(\psi^u(x) - \psi^s(x))$. It is easily seen that $l(\gamma_\Delta^u(x)) \approx p(\gamma_\Delta^u(x))/(\cos\varphi(x)\cos\psi^u(x))$ and $l(\gamma_\Delta^s(x)) \approx p(\gamma_\Delta^s(x))/(\cos\varphi(x)\cos\psi^s(x))$, hence, by (2.4), $\nu(\Delta) \approx p(\gamma_\Delta^u(x))\, p(\gamma_\Delta^s(x))(B^u(x) + B^s(x))c_\nu$.

Returning to our parallelogram U, making up the integral sum $\nu(U) = \Sigma\nu(\Delta_{ij})$, and passing to the limit, as in the usual definition of integral, we obtain

$$\nu(U) = c_\nu \int_U (B^u(x) + B^s(x))\, d\nu_U(x), \tag{3.3}$$

where $d\nu_U(x)$ denotes the measure on U which is the direct product of the measure $p(\cdot)$ on the sets $\gamma_U^u(x)$ and $\gamma_U^s(x)$. Since addition on the LUM and LSM is absolutely continuous [13], [29], almost every point $x \in U$ is a density point of the set $\gamma_U^{u,s}(x)$ on the LUM $\gamma^u(x)$ (the LSM $\gamma^s(x)$), and thus the Jacobian of the canonical isomorphism of this set onto its projection Γ_U^u (Γ_U^s) on the corresponding coordinate axis is defined at it. We denote this Jacobian by $J^u(x)$ ($J^s(x)$). Then (3.3) takes the form

$$\nu(U) = c_\nu \int_{\Gamma_U^u} dp(y) \int_{\Gamma_U^s} dp(z)\, (B^u(x) + B^s(x))\, J^u(x)\, J^s(x). \tag{3.4}$$

Here $y \in \Gamma_U^u$ and $z \in \Gamma_U^s$ are points such that $[y, z] = x$. It is natural to regard y and z as the *coordinates* of x on the axes $\gamma^u(x_0)$ and $\gamma^s(x_0)$. It is important to note that all functions appearing under the integral sign in (3.4) are determined by x and do not depend on the choice of the parallelogram U.

3.2. Weak homogeneity of parallelograms.

Formula (3.4) represents the parallelogram U as a direct product in the metrical sense (and not just the topological sense, as in (3.2)). However, at present it is only convenient if the integrand is almost constant. For such parallelograms we introduce the notion of homogeneity.

First we fix constants $\alpha_0 < 1$ and $C_0 > 0$.

Definition 3.1. A parallelogram U is called *weakly n-homogeneous* (where $n \geqslant 0$ is an integer) if the following four conditions hold:

(01) $| B^u(x)/B^u(y) - 1 | \leqslant C_0\alpha_0^n$;
(02) $| B^s(x)/B^s(y) - 1 | \leqslant C_0\alpha_0^n$;
(03) $| J^u(x) - 1 | \leqslant C_0\alpha_0^n$;
(04) $| J^s(x) - 1 | \leqslant C_0\alpha_0^n$

for any points $x, y \in U_0$ and any point $x_0 \in U_0$ fixing the coordinate axes $\gamma^u(x_0)$ and $\gamma^s(x_0)$.

The next lemma easily follows from (3.4).

***Lemma* 3.2.** *The measure of subparallelograms U of weakly n-homogeneous parallelograms U_0 can be approximated by the quantity*

$$\nu_a(U) = c_\nu p(\Gamma_U^u) p(\Gamma_U^s)(B^u(x_0) + B^s(x_0)). \tag{3.5}$$

More precisely,

$$| \nu_a(U)/\nu(U) - 1 | \leqslant C_1\alpha_0^n, \tag{3.6}$$

where $C_1 = C_1(C_0, \alpha_0)$.

We derive one consequence from the fomulae (3.5)–(3.6). A parallelogram $U' \subset U$ is said to be *u-inscribed* (*s-inscribed*) in U if $\gamma^u_{U'}(x) = \gamma^u_U(x)$ (respectively, $\gamma^s_{U'}(x) = \gamma^s_U(x)$) for any point $x \in U'$. Let U' be an arbitrary parallelogram which is u-inscribed in U, and let U'' be an arbitrary parallelogram s-inscribed in U. We also assume that U is weakly n-homogeneous. Then (3.5), (3.6) imply the inequality

$$|\nu\ (U''/U)/\nu\ (U''/U') - 1\ | \leqslant C_2\alpha_0^n, \tag{3.7}$$

where $C_2 = C_2(C_1, \alpha_0)$ (here $\nu(A/B)$ denotes the conditional measure).

3.3. Homogeneous LUM and LSM.

It is necessary to study the construction of the above-defined weakly n-homogeneous parallelograms. Note that for weak n-homogeneity it is not sufficient that the parallelogram has small diameter. For example, if two parts of a parallelogram U lie on *different* sides of some curve $\gamma \subset R_{0,k}$ for a small value of k, then the quantity $B^u(x)$ has a jump when crossing γ which does not vanish when diam $U \to 0$. Therefore, for weak n-homogeneity we must additionally require that the images of the parallelogram U, both in the past and the future, do not experience discontinuities for a sufficiently long time. Moreover, it is necessary that the functions $B^{u,s}(x)$ do not strongly oscillate inside U and its images T^iU for small (in absolute value) values of i. These functions have singularities on the set $S_0 \subset \partial M$, and in the presence of break points on the boundary ∂Q also at finitely many "singular" points on V_0, which were mentioned in §2 and are described in more detail in Appendix 1. Therefore we partition a neighbourhood of S_0 and a neighbourhood V_* of the set of these "singular" points into countably many subdomains, contracting towards S_0 and the "singular" points, and inside which the functions $B^{u,s}(x)$ do not strongly oscillate.

We fix a certain $\nu > 1$ and integer $n_0 \geqslant 1$. We draw in $\overline{M}$ the countably many segments given by the equations $\varphi = \pi/2 - n^{-\nu}$ and $\varphi = -\pi/2 + n^{-\nu}$ for all integers $n \geqslant n_0$. We denote the set of these segments by $\mathscr{D}_0$. We consider also the images and inverse images of the points on these segments which fall in the set V_*, up to the time when they leave this set. These points form the set $\mathscr{D}_1 \subset V_*$, consisting of at most countably many curves. In other words, $\mathscr{D}_1$ consists of the images of the segments in $\mathscr{D}_0$ which "get stuck" for some time in small neighbourhoods of break points of ∂Q. We will give a more precise definition of the set $\mathscr{D}_1$ in Appendix 1 (§A1.3). We put $\mathscr{D} = \mathscr{D}_0 \cup \mathscr{D}_1$.

Definition 3.3. We call a segment LUM γ^u (LSM γ^s) *homogeneous* (for short, HLUM and HLSM) if its images $T^{-n}\gamma^u$ ($T^n\gamma^s$) for all $n \geqslant 0$ (that is, when "moving in the direction of contraction") do not intersect the segments and curves in the set $\mathscr{D}$ (but their end-points may fall on them) and, moreover, T (T^{-1}) is smooth on γ^u (γ^s) (that is, one iteration in the "direction of expansion" will not lead to discontinuity).

In other words, the construction of an HLUM (HLSM) can be realized by breaking up an arbitrary LUM (LSM) at the points whose images in the past (future) fall in $\mathcal{D}$. The latter will therefore be called *subdividing curves*, and the segments in $\mathcal{D}_0$ *subdividing segments*.

Definition 3.4. Let $n \geqslant 1$ be an integer. We call an LUM γ^u (LSM γ^s) *n-homogeneous* if its image $T^n\gamma^u$ ($T^{-n}\gamma^s$) is an HLUM (HLSM).

Of course, the property of being n-homogeneous is stronger than that of being simply homogeneous.

Definition 3.5. We call a parallelogram U *n-homogeneous* if for any point x of it the set $\gamma^u_U(x)$ ($\gamma^s_U(x)$) lies entirely on one n-homogeneous LUM (LSM).

Note that the closure in $\overline{M}$ of any n-homogeneous parallelogram is also an n-homogeneous parallelogram.

Theorem 3.6. *The parallelograms that are n-homogeneous are also weakly n-homogeneous, for a corresponding choice of constants in Definitions* 3.1 *and* 3.3 – 3.5. *More precisely, the constants* ν *and* n_0 *can be chosen arbitrarily, while* α_0 *and* C_0 *will depend on them.*

The meaning of this theorem is sufficiently transparent, but its proof is lengthy and is for a large part technical in nature, hence we transfer it to Appendix 1.

We must note that the n-homogeneous LUM and LSM have another important property. We denote by $\Lambda^u_i(x)$, for $i \geqslant 1$, the local coefficient of contraction of the LUM $\gamma^u(x)$ at the point x under the action of T^{-i}.

Lemma 3.7. *If the* LUM γ^u *is n-homogeneous, then for any* $x, y \in \gamma^u$ *and* $i \geqslant 1$ *the estimate* $|\Lambda^u_i(x)/\Lambda^u_i(y) - 1| \leqslant C_4\alpha_0^n$ *holds, where* $C_4 > 0$ *is a constant. A similar estimate holds for n-homogeneous* LSM.

The proof of Lemma 3.7 will also be given in Appendix 1.

Theorem 3.6 indicates a way of constructing n-homogeneous parallelograms, but leaves open the question of the existence of n-homogeneous LUM and LSM. In fact, if the subdividing points densely filled each LUM (LSM), then not a single HLUM (HLSM) could be chosen at them.

Theorem 3.8 (existence). *For almost any point* $x \in M$ *there is an* HLUM (HLSM) *containing* x *in its interior.*

We denote by $\gamma^{0u}(x)$ ($\gamma^{0s}(x)$) the maximal smooth segment HLUM (HLSM) passing through x.

Corollary 3.9. *For any* $n \geqslant 1$ *and almost any point* x *there is an n-homogeneous* LUM (LSM) *passing through* x.

Theorem 3.8 can be substantially refined by estimating the distribution of the lengths on HLUM and HLSM in M. In fact, let $r^{0u}(x)$ ($r^{0s}(x)$) be the distance from the point $x \in M$ to the nearest end-point of the curve $\gamma^{0u}(x)$

($\gamma^{0s}(x)$) in the p-metric (if this curve does not exist, we set this distance equal to zero).

***Theorem* 3.10.** *$\nu\{x \in M : r^{0u}(x) \leqslant \varepsilon\} \leqslant C_5\varepsilon^{\theta}$, where $C_5 = C_5(\nu, n_0) > 0$ and $\theta = \theta(\nu, n_0) > 0$. A similar estimate holds for $r^{0s}(x)$.*

It is useful to note the following completion to this.

***Proposition* 3.11.** *For ν-almost any $x \in M$ the subdividing points on the* LUM *$\gamma^u(x)$ (LSM $\gamma^s(x)$) partitioning this curve into individual* HLUM (HLSM) *can only occur at the end-points of this curve, and do not have limit points inside it.*

The proofs of assertions 3.8 – 3.11 are very similar to the method of constructing LUM and LSM described in [13], [29]. They are carried out in Appendix 2.

With this we finish the analysis of the properties of parallelograms. The remaining part of §3 is devoted to the dynamics of individual HLUM and HLSM.

3.4. Rate of expansion of HLUM and HLSM.

The aim of this and the next subsection is to study the dynamics of HLUM and HLSM when they move under the action of T^n "in the direction of expansion". The local coefficient of expansion after n steps grows exponentially with n (2.2), but the global picture is much more complicated, since expansion is accompanied by discontinuities and subdivisions. As a result, for each $n \geqslant 1$ the image $T^n\gamma^u$ ($T^{-n}\gamma^s$) consists of finitely or countably many HLUM (HLSM), among which there may be arbitrarily short ones. We call these HLUM (HLSM) the *homogeneous components* of the image $T^n\gamma^u$ ($T^{-n}\gamma^s$). In this subsection we show that, nevertheless, the average length of these components grows exponentially with n, until it becomes a quantity of order one. In other words, the process of expansion dominates over the counterprocesses of discontinuity and subdivision. It is important to stress that here the discussion is about *any* HLUM (HLSUM), and not *almost any*, as before.

In the subsequent exposition only HLUM figure, but similar results are valid also for HLSM, with n replaced by $-n$.

Let γ^u be an arbitrary HLUM of length $p(\gamma^u) = p$, and let $D > 0$. For each $n \geqslant 0$ we partition γ^u into subsegments $\gamma^u_{i,n}$, $i = 1, 2, \ldots$, which are transformed under the action of T^n into homogeneous components of length $\geqslant D$. We consider the quantity

$$p_{D,\gamma^u}(N) = p\{\bigcup_{i,\, n\leqslant N} \gamma^u_{i,n}\}/p(\gamma^u).$$

In other words, $p_{D,\gamma^u}(N)$ is the relative fraction of points on γ^u whose images during the first N steps fall *at least once* in homogeneous components of length $\geqslant D$.

Theorem 3.12 (rate of expansion). *There are a* $D > 0$ *and a function* $\beta(c)$, $\beta(c) \to 0$ *as* $c \to \infty$, *independent of* γ^u *and* p, *and such that for any* $c > 0$ *we have* $p_{D,\gamma^u}(-[c \ln p]) \geqslant 1 - \beta(c)$.

Theorem 3.2 means that during the first N iterations of T, where $N = -[c \ln p(\gamma^u)]$, the majority of points on γ^u fall at least once in long homogeneous components (of length $\geqslant D$). The logarithmic dependence of N on $p(\gamma^u)$ means that the expansion of the average length of homogeneous components of the image $T^n\gamma^u$ is of exponential rate. This is, in essence, the key property of the hyperbolic systems under consideration, determining a fast decay of correlations.

In the simpler case when subdivisions do not have to be taken into account (that is, instead of HLUM we consider LUM as components of the set $T^n\gamma^u$), a similar (and in a certain respect, stronger) assertion has already been proved in [24]. Taking subdivisions into account somewhat complicates the proof of the theorem. We give it in Appendix 3.

3.5. Properties of transitivity for HLUM and HLSM.
In this subsection we continue the study of the dynamics of HLUM and HLSM when they move under the action of T^n "in the direction of expansion". Here we assume that the initial HLUM γ^u already has length of order one. In this case its images already "nowhere" expand, and therefore the homogeneous components of these images start to fill the space M. General considerations, appealing to the mixing property of the transformation T [13], predict that for large n the homogeneous components of the set $T^n\gamma^u$ are uniformly distributed on M. We will need a much weaker property, consisting, roughly speaking, of the fact that the density of filling M by these components is asymptotically bounded away from zero as $n \to \infty$. It is this property that we call *transitivity*. Below we give strict formulations.

First of all we introduce a number of new notions for describing the metrical structure of parallelograms. They aid in better formulating and more clearly demonstrating the property of transitivity.

A *quadrangle* is a domain $K \subset \overline{M}$ bounded by a pair of LUM and a pair of LSM, in alternation. We call the bounding LUM (LSM) *u-faces* (*s-faces*). For each parallelogram U we can choose a minimal quadrangle $K(U)$ containing it. We call $K(U)$ the *carrier* of U, and the s- and u-faces of U are taken to be those of its carrier.

Further, we say that the LUM γ^u (LSM γ^s) is *stretched* on the quadrangle K (parallelogram U) if its end-points lie *precisely* on the s-faces (u-faces) of this quadrangle (parallelogram). It is easily verified that if the parallelogram is n-homogeneous, then its s- and u-faces, as well as all HLUM and HLSM streteched on it, are n-homogeneous. We say that an n-homogeneous parallelogram ($n \geqslant 0$) is *maximal* if it intersects all HLUM and HLSM stretched on it. Any n-homogeneous parallelogram can be completed to a maximal one while preserving n-homogeneity and not splitting its boundary.

We also say that the quantity $p(A \cap \gamma)/p(\gamma)$ is the *density of the set* $A \subset M$ *on the curve* γ.

We now turn to the formulation of the property of being transitive. We consider an arbitrary HLUM γ^u of length $p(\gamma^u) = p$ and an arbitrary 0-homogeneous maximal parallelogram U_0 of non-zero measure. For each $n \geqslant 1$ we choose inside γ^u subsegments $\hat{\gamma}^u_{i,n}$, $i \geqslant 1$, whose images $T^n\hat{\gamma}^u_{i,n}$ are stretched on the parallelogram U_0 (if there are such), and we consider the quantity

$$p_n(\gamma^u, U_0) = p(\bigcup_i \hat{\gamma}^u_{i,n})/p(\gamma^u).$$

Theorem **3.13** (transitivity). *Under the above-described conditions,* $p_n(\gamma^u, U_0) \geqslant \delta_0(p, U_0) > 0$ *for all* $n \geqslant n_0(p, U_0)$, *where* δ_0 *and* n_0 *depend on* U_0 *and on the length* p *of the initial curve* γ^u, *but not on its position.*

Theorem 1.3 means that if the initial curve γ^u is sufficiently long, then after sufficiently many iterations its image will always contain a substantial part consisting of HLUM stretched on the parallelogram U_0. From the fact that U_0 is arbitrary it follows that for large n the components of the image of the curve $T^n\gamma^u$ actually "unravel" on M with density bounded away from zero. (In view of the mixing property of T there must clearly exist some relation between $\delta_0(p, U_0)$ and the measure $\nu(U_0)$; however, we will not consider this question here.)

We again stress that, as in §3.4, the discussion is about *any* HLUM γ^u. The dual assertion is true for any HLSM.

The proof of Theorem 3.13 is very intuitive, and is based on a number of assertions of independent interest, hence we give it here and do not transfer it to the Appendices. The idea of the proof is also applicable to non-billiard hyperbolic maps with singularities.

Lemma **3.14.** *Let* γ *be an arbitrary increasing curve in* $\overline{M}$. *Then through almost every* (*in the sense of the p-measure*) *point* $x \in \gamma$ *there passes an* LSM $\gamma^s(x)$.

The assertion of the lemma is, in essence, an analogue of the main theorem of the theory of hyperbolic billiards. At present two versions of its proof are known. The first was worked out in [13], [5], [6], [21], and the second, simpler, one in [15], [31], [25]. Lemma 3.14 follows immediately from the first version, while it follows after minor and obvious modifications in the proof, not given here, from the second version.

Assertions 3.11 and 3.14 immediately imply the following result.

Corollary **3.15.** *Let* γ *be an arbitrary increasing curve in* $\overline{M}$. *Then through almost every point* $x \in \gamma$ *there passes an* HLSM $\gamma^{0s}(x)$.

Remark 3.16. The absolute continuity of addition on LUM and LSM [13], [29] implies that the measure of the union of all HLSM involved in

Corollary 3.15 is positive. Hence there is a $\rho_0 = \rho_0(\gamma) > 0$ such that if the HLSM are drawn only through the points in $\{x \in \gamma : r^{0s}(x) \geqslant \rho_0\}$ (that is, only those HLSM are taken for which the distance from the end-points to the point of intersection with γ is not less than ρ_0), then the measure of their union is also positive. Moreover, they "cut out" on γ a set whose p-measure tends to $p(\gamma)$ as $\rho_0 \to 0$.

Using assertions 3.15 and 3.16 it is easy to construct in a neighbourhood of an arbitrary HLUM γ^u, and for arbitrary $\varepsilon_1 > 0$ and $d \in (0, 1/10)$, a parallelogram U_1 with the following set of properties:

1. U_1 is 0-homogeneous and maximal;
2. γ^u intersects both s-faces of U_1, and the points of intersection have distance larger than $dp(\gamma^u)$ to the end-points of γ^u;
3. for each HLSM γ^s stretched on U_1 the point of intersection $\gamma^s \cap \gamma^u$ has distance larger than $dp(\gamma^s)$ from the end-points of γ^s (in other words, the curve γ^u passes somewhere in the central part of the quadrangle $K(U_1)$ and does not come too close to its u-faces);
4. the density of the parallelogram U_1 on every HLSM stretched on it is at least $1 - \varepsilon_1$.

We call a parallelogram U_1 having the properties $1-4$ a *dense part (with parameters* ε_1 *and* d*) of the* HLUM γ^u.

***Lemma* 3.17** (on finite collections of dense parts). *For any $p > 0$, $\varepsilon_1 > 0$, and $d \in (0, 1/10)$ we can choose in $\overline{M}$ a finite collection of parallelograms such that for each* HLUM *of length $\geqslant p$ one of these parallelograms is a dense part with parameters ε_1 and d.*

For the proof we introduce on the set of all HLUM a metric given by the maximum of the distances from each point on one HLUM to the nearest point on another HLUM, and conversely. By the compactness of the space, for any $p > 0$ the subset of all HLUM of lengths $\geqslant p$ is compact. It is also easily seen that each parallelogram U_1 serves as a dense part for some open subset of HLUM of length $\geqslant p$ (possibly empty), which implies the lemma.

We fix a $d \in (0, 1/10)$, a sufficently small $\varepsilon_1 > 0$, and a dense part with parameters ε_1, d for the curve γ^u in the conditions of Theorem 3.13.

For the moment we forget about the parallelogram U_1, and return to the parallelogram U_0 in the conditions of Theorem 3.13. The absolute continuity of addition on LUM and LSM implies that almost any point $x \in U_0$ is a density point of the measurable set $\gamma^s_{U_0}(x)$ on the curve $\gamma^s(x)$. The definition of density point implies that for any $\varepsilon_2 > 0$ there are an $l_0 > 0$ and a subset $\tilde{U}_0 \subset U_0$ of non-zero measure such that on any HLSM intersecting $\tilde{U}_0$ and of p-length less than l_0 the parallelogram U_0 has density $\geqslant 1 - \varepsilon_2$. Here, of course, the set $\tilde{U}_0$ need not be a parallelogram.

We fix a sufficiently small $\varepsilon_2 > 0$ and a corresponding set $\tilde{U}_0$.

We consider the set $T^n U_1$, for an arbitrary $n \geqslant 1$. It consists of finitely many 0-homogeneous parellelograms. We denote by $U_{1,n}, \ldots, U_{k(n),n}$ those parallelograms that intersect $\tilde{U}_0$. The mixing property of T and Lemma 3.17 imply that for all $n \geqslant n'(p, \varepsilon_1, d, U_0, \varepsilon_2)$ the inequality

$$\sum_{i=1}^{k(n)} \nu(U_{i,n} \cap U_0) > q_1 \tag{3.8}$$

holds, where $q_1 = q_1(p, \varepsilon_1, d, U_0, \varepsilon_2) > 0$.

Under the action of the canonical isomorphism each parellelogram $U_{i,n}$, $1 \leqslant i \leqslant k(n)$, is projected onto one of the homogeneous components $\gamma_{i,n}^u$ of the image $T^n\gamma^u$; moreover, *distinct* parallelograms $U_{i,n}$ project to *distinct* homogeneous components $\gamma_{i,n}^u$. (Properly speaking, on account of this correspondence we will also study in detail the properties of homogeneous components of the images of LUM and LSM; in §4 we will derive from it the required properties of the parallelograms themselves.)

***Lemma* 3.18.** *Each homogeneous component $\gamma_{i,n}^u$, $1 \leqslant i \leqslant k(n)$, intersects both s-faces of the parallelogram U_0 provided that n is sufficiently large* ($n \geqslant n''(U_0, \varepsilon_2)$) and ε_1, ε_2 *are sufficiently small.*

The proof rests on a geometrical construction, which is illustrated in Fig. 6. We choose a point $y \in U_{i,n} \cap \tilde{U}_0$ and, correspondingly, the HLSM $\gamma_1^s = \gamma^s(T^{-n}y) \cap K(U_1)$. For sufficiently large n its image $\gamma_0^s = T^n\gamma_1^s$ has length $< l_0$, and hence the density of the parallelogram U_0 on γ_0^s is not less than $1-\varepsilon_2$. This and properties 3, 4 of the dense part U_1 imply that for sufficiently small ε_1, ε_2 the curve γ_0^s contains "sufficiently many" points of the set $T^nU_1 \cap U_0$, while the point $\gamma_{i,n}^u \cap \gamma_0^s$ lies between some pair of points $y_1, y_2 \in T^nU_1 \cap U_0$. We consider the quadrangle K^* bounded by the LUM $\gamma^u(y_1)$ and $\gamma^u(y_2)$ and the s-faces of U_0. It is easily seen that its inverse images $T^{-i}K^*$ for all $i = 1, 2, \dots, n$ do not intersect the discontinuity curves and the subdividing curves, which implies the lemma.

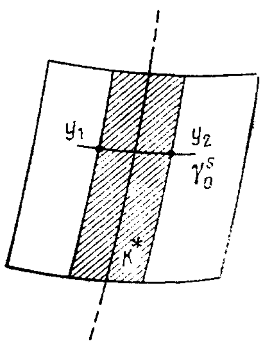

Fig. 6

Lemma 3.18 shows that on each homogeneous component $\gamma^u_{i,n}$ of the image $T^n\gamma^u$ there is a subsegment $\hat{\gamma}^u_{i,n}$ which is stretched on U_0. It is obvious that $p(\hat{\gamma}^u_{i,n}) \geqslant \text{const}(U_0)\cdot p(\gamma^u_{i,n})$, hence (Lemma 3.7)

$$p(T^{-n}\hat{\gamma}^u_{i,n}) \geqslant C_4^{-1}\,\text{const}(U_0)\cdot p(T^{-n}\gamma^u_{i,n}).$$

This and (3.8) imply the estimate

$$\sum_{i=1}^{k(n)} p(T^{-n}\hat{\gamma}^u_{i,n}) \geqslant q_2 \sum_{i=1}^{k(n)} \nu(T^{-n}(U_{i,n}\cap U_0)) > q_1 q_2, \tag{3.9}$$

where the quantity $q_2 = q_2(p, \varepsilon_1, d, U_0, \varepsilon_2) > 0$ appears by applying formula (3.4) for computing the measure of the parallelograms $T^{-n}(U_{i,n} \cap U_0)$, taking into account that U_1 is 0-homogeneous and Lemma 3.17. Inequality (3.9) holds for all $n \geqslant \max\{n', n''\}$, which proves Theorem 3.13.

Successive application of Theorem 3.12 and 3.13 leads to the following general assertion.

***Theorem* 3.19.** *Suppose that the* HLUM γ^u *and parallelogram* U_0 *satisfy the conditions of Theorem* 3.13. *Then* $p_n(\gamma^u, U_0) > \delta_1(U_0)$ *for all* $n \geqslant -[C^{(1)}\ln p]+n_1(U_0)$, *where the quantities* $\delta_1(U_0)$ *and* $n_1(U_0)$ *depend only on* U_0, *and* $C^{(1)} = C^{(1)}(Q) > 0$.

§4. The Markov lattice: global construction

The Markov lattice is an instrument whose application allows us to derive Theorems 1.1 and 1.2 relatively quickly in §5, §6. Clearly, it is also useful in a subsequent, more detailed study of the statistical properties of hyperbolic billiards, hence we will try to describe it in maximally general and, at the same time, detailed form.

4.1. Definition and basic properties.

A Markov lattice is given by a pair of integer-valued parameters $N > n > 0$. We immediately note that N is the length of the time interval on which we will approximate our process $\{X_i\}$ by a process of Markov type (as explained in §1). Formally we will not impose further restrictions on N and n, but it will be useful to keep in mind that we are interested in the case when $N, n \to \infty$ and $n \sim N^\gamma$ for some $\gamma < 1$.

By a *Markov lattice with parameters n, N* we will mean a finite partition of the space $M : \mathcal{R}_{n,N} = \{V_0, V_1, \ldots, V_I\}$, where $I = I(n, N)$, $\nu(V_i \cap V_j) = 0$ for $i \neq j$, and $\bigcup \overline{V}_i = \overline{M}$, having the following four properties:

***Property* ML1** (dimensions). diam $V_i \leqslant e^{-n}$ *for all* $i \geqslant 1$ (*the set* V_0 *is "special", it does not participate in this estimate*).

***Property* ML2** (measure of the remainder). $\nu(V_0) \leqslant Ne^{-n}$.

Property **ML3** (Markov approximation). *For any integers $k > l > 1$ and $1 \leqslant i_1 < i_2 < \ldots < i_k \leqslant N$, as well as for indices $j_1, j_2, \ldots, j_k$ taking values from 1 to $I = I(n, N)$, the following holds:*

$$(4.1)\qquad \nu(T^{i_1}V_{j_1} \cap T^{i_2}V_{j_2} \cap \cdots \cap T^{i_{l-1}}V_{j_{l-1}}/T^{i_l}V_{j_l} \cap \cdots \cap T^{i_k}V_{j_k}) = \nu(T^{i_1}V_{j_1} \cap \cdots \cap T^{i_{l-1}}V_{j_{l-1}}/T^{i_l}V_{j_l})(1+\Delta),$$

where $|\Delta| \leqslant C_5\alpha_0^n$ for some $C_5 = C_5(\alpha_0)$ (the quantity α_0 was introduced in §3.2).

Relation (4.1) denotes, up to a factor $1+\Delta$, the Markov property. It is easily proved that (4.1) holds also for $N \geqslant i_1 > i_2 > \ldots > i_k \geqslant 1$.

For convenience of notation we denote by $\mathcal{J}$ the set of indices $\{0, 1, \ldots, I\}$, where $I = I(n, N)$. For any $k \geqslant 1$ and $i \in \mathcal{J}$ we choose a subset $R_i(k) \subset \mathcal{J}$ such that $j \in R_i(k)$ if and only if

$$(4.2)\qquad \nu(T^kV_i \cap V_j) \geqslant \beta_0\nu(V_i)\,\nu(V_j),$$

where $\beta_0 = \beta_0(Q) > 0$. Moreover, for $k \geqslant 1$ we define the subset $R(k) \subset \mathcal{J}$ such that $i \in R(k)$ if and only if

$$(4.3)\qquad \sum_{j\in R_i(k)} \nu(V_j) > 1 - e^{-n}.$$

Property **ML4** (regularity). *For each $k \geqslant D_0 n$ we have*

$$(4.4)\qquad \sum_{i\in R(k)} \nu(V_i) > 1 - Ne^{-n},$$

where $D_0 = D_0(Q) > 0$.

The constants β_0 and D_0 are defined in §4.3, but their numerical values do not play a role in the sequel.

The general idea of the relations (4.3) and (4.4) consists in the fact that inequality (4.2), which guarantees a certain degree of mixing after k steps, holds for "the overwhelming majority" of pairs of indices i, j. Below, in Theorem 4.1, this inequality will allow us to prove an analogue of the strong mixing condition in the sense of Ibragimov [9]. The construction of a Markov lattice will be given in §4.2–§4.4.

Theorem **4.1** (convergence towards equilibrium). *Let $k > l \geqslant 1$ be integers and $1 \leqslant i_1 < i_2 < \ldots < i_k \leqslant N$. Then the set $\mathcal{J}^{k-l}$ contains a subset $R_* = R_*(i_1, \ldots, i_k)$ of tuples of $(k-l)$ indices such that:*

a) *if $(j_{l+1}, \ldots, j_k) \in R_*$, then*

$$(4.5)\qquad \sum_{j_1,\ldots,j_l=0}^{I} |\nu(T^{i_1}V_{j_1} \cap \cdots \cap T^{i_l}V_{j_l}/T^{i_{l+1}}V_{j_{l+1}} \cap \cdots \cap T^{i_k}V_{j_k}) - \nu(T^{i_1}V_{j_1} \cap \cdots \cap T^{i_l}V_{j_l})| \leqslant \Delta;$$

b) *the following estimate holds*:

$$(4.6) \qquad \sum_{(j_{l+1},\dots,j_k)\in R_*} \nu(T^{i_{l+1}}V_{j_{l+1}} \cap \dots \cap T^{i_k}V_{j_k}) \geqslant 1-\Delta,$$

where Δ *can be taken to be* $\Delta = \max\{N^2e^{-n/2},\ CN\alpha_0^n,\ (1-\beta_0)^{[L/2]}\}$ *for* $L = [(i_{l+1}-i_l)/(D_0 n)]$.

Remark 4.2. It is easily verified that the assertion of the theorem is equally valid for $N \geqslant i_1 > i_2 > \dots > i_k \geqslant 1$.

Part a) of Theorem 4.1 means that the conditional distributions on the sets $T^{i_1}V_{j_1} \cap \dots \cap T^{i_l}V_{j_l}$ converge sufficiently fast to the *unconditional* one (more precisely, exponentially fast with the growth of the length of the time interval $|i_{l+1}-i_l|$ between the "past" and the "future", admittedly, as long as this interval does not become a quantity of order n^2). By part b) of Theorem 4.1, this convergence holds for "the overwhelming majority" of possible conditions $T^{i_{l+1}}V_{j_{l+1}} \cap \dots \cap T^{i_k}V_{j_k}$.

We will complete the proof of Theorem 4.1 in §4.1. It uses fairly standard methods from the theory of Markov chains, hence in a number of places we leave out the details. First, Property ML2 implies that a tuple of indices $(j_1, \dots, j_k)$ including at least one zero can be neglected: these do not have an influence on the truth of inequalities (4.5) and (4.6). By Property ML3 we can apply Markov approximation to the remaining tuples of indices (not containing zero). Property ML3 also allows us to immediately reduce Theorem 4.1 to the case $k = l+1$, that is, when only the single set $T^{i_{l+1}}V_{j_{l+1}}$ figures in formula (4.5). It is somewhat more complicated to reduce the theorem to the case $l = 1$, but even this can be done (by successively getting rid of the indices $j_1, j_2, \dots, j_{l-1}$ in (4.5)). We only indicate the first step of this procedure:

$$\begin{aligned}\nu(T^{i_1}V_{j_1} \cap \dots \cap T^{i_l}V_{j_l}/T^{i_{l+1}}V_{j_{l+1}}) - \nu(T^{i_1}V_{j_1} \cap \dots \cap T^{i_l}V_{j_l}) =\\ = \nu(T^{i_1}V_{j_1}/T^{i_2}V_{j_2} \cap \dots \cap T^{i_{l+1}}V_{j_{l+1}}) \cdot \nu(T^{i_2}V_{j_2} \cap \dots \cap T^{i_l}V_{j_l}/T^{i_{l+1}}V_{j_{l+1}}) -\\ - \nu(T^{i_1}V_{j_1}/T^{i_2}V_{j_2} \cap \dots \cap T^{i_l}V_{j_l}) \cdot \nu(T^{i_2}V_{j_2} \cap \dots \cap T^{i_l}V_{j_l}).\end{aligned}$$

Further, to the two conditional measures on the right-hand side (with "long" conditions) we apply formula (4.1), and subsequently sum over j_1.

It remains to prove Theorem 4.1 for $l = 1$ and $k = 2$. We may assume that $L \geqslant 2$, otherwise the theorem is trivial, since the set R_* can be chosen to be empty. For convenience of notation we put $t_0 = i_1$, $t_L = i_2$ $(= i_k)$, and choose arbitrary $t_1 < t_2 < \dots < t_{L-1}$ such that $t_0 < t_1$, $t_{L-1} < t_L$ and $\min\{t_1-t_0, t_2-t_1, \dots, t_L-t_{L-1}\} \geqslant D_0 n$ (that is, we fix $L-1$ intermediate moments of time in the interval $[t_0, t_L]$, with pairwise distances $\geqslant D_0 n$). After this we may write

$$\begin{aligned}(4.7) \qquad \nu(T^{t_0}V_{j_0}/T^{t_L}V_{j_L}) = \sum_{j_1,\dots,j_{L-1}=0}^{I} \nu(T^{t_{L-1}}V_{j_{L-1}}/T^{t_L}V_{j_L}) \times\\ \times \nu(T^{t_{L-2}}V_{j_{L-2}}/T^{t_{L-1}}V_{j_{L-1}} \cap T^{t_L}V_{j_L}) \times \dots \times \nu(T^{t_0}V_{j_0}/T^{t_1}V_{j_1} \cap \dots \cap T^{t_L}V_{j_L}).\end{aligned}$$

It is easily proved that the tuples $j_1, \ldots, j_{L-1}$ including at least one zero contribute to the sum in (4.7) by at most $e^{-n/2}$. We apply Markov approximation (Property ML3) to the remaining tuples, reducing (4.7) to the sum

$$\sum_{j_1,\ldots,j_{L-1}=0}^{I} \prod_{l=1}^{L} \nu(T^{t_{l-1}}V_{j_{l-1}}/T^{t_l}V_{j_l}). \tag{4.8}$$

For convenience we introduce new, "probabilistic", notations. We put $\pi_{ij}^{(l)} = \nu(T^{t_{l-1}}V_j/T^{t_l}V_i)$ for $1 \leqslant l \leqslant L$ and $0 \leqslant i, j \leqslant I$. Then the matrices $\Pi^{(l)} = \| \pi_{ij}^{(l)} \|$ are stochastic for any l and have a common stationary vector $P = \|p_i\|$, $p_i = \nu(V_i)$. For $1 \leqslant l \leqslant L$ we denote the product of the matrices $\Pi^{(1)} \cdot \Pi^{(2)} \cdot \ldots \cdot \Pi^{(l)}$ by $\Pi^{(1,l)} = \| \pi_{ij}^{(1,l)} \|$. Also, we denote by $P_i^{(1,l)}$ the i-th row of the matrix $\Pi^{(1,L)}$. Then the sum (4.8) is nothing but the entry $\pi_{j_0 j_L}^{(1,L)}$ of the matrix $\Pi^{(1,L)}$. Hence the required inequality (4.5) (for $l = 1$, $k = 2$) can be written in the form

$$\sum_{j=0}^{I} |\pi_{ij}^{(1,L)} - p_j| = 2\mathrm{Var}(P_i^{(1,L)}, P) \leqslant \Delta, \tag{4.9}$$

where $\mathrm{Var}(\cdot, \cdot)$ is the distance in variation between the two probability vectors. Smallness of the left-hand side of (4.9) means rapid convergence to equilibrium in the non-stationary Markov chain with transition matrix $\Pi^{(l)}$, $l = 1, 2, \ldots, L$. The following technical lemma is the first step in proving (4.9).

***Lemma* 4.3.** *For any $l = 1, 2, \ldots, L$ and $i \in \mathcal{I}$ we have*

$$\mathrm{Var}(P_i^{(1,l)}, P) \leqslant (1 - \beta_0)\,\mathrm{Var}(P_i^{(1,l-1)}, P) + \Delta',$$

where

$$\Delta' = 6\Big(\sum_{j \notin R(t_l - t_{l-1})} \pi_{ij}^{(1,l-1)} + e^{-n}\Big).$$

We recall that $R(\cdot)$ denotes the subset of $\mathcal{I}$ introduced in ML4.

Proof of the lemma. For shortness of notation we put $\tau_l = t_l - t_{l-1}$. First we note that by Property ML4, for all $i \in R(\tau_l)$,

$$\sum_{j \in R_i(\tau_l)} \pi_{ij}^{(l)} \geqslant 1 - \beta_0 e^{-n}. \tag{4.10}$$

Further, it is easily proved that

$$\begin{aligned} \mathrm{Var}(P_i^{(1,l)}, P) = {} & \sum_j{}^{+} \sum_{k \in R(\tau_l)} (\pi_{ik}^{(1,l-1)} - p_k)(\pi_{kj}^{(l)} - \beta_0 p_j) + \\ & + \beta_0 \sum_j{}^{+} \sum_{k \in R(\tau_l)} (\pi_{ik}^{(1,l-1)} - p_k)\,p_j + \sum_j{}^{+} \sum_{k \notin R(\tau_l)} (\pi_{ik}^{(1,l-1)} - p_k)\,\pi_{kj}^{(l)}, \end{aligned} \tag{4.11}$$

where Σ_j^{+} denotes the sum over those $j \in \mathcal{I}$ such that $\pi_{ij}^{(1,l)} > p_j$. It is easily verified that the second and third terms in (4.11) do not exceed the quantity

$\Delta'/6$ from the lemma. The first term can be majorized by the sum

$$(4.12)\qquad \sum_{k\in R(\tau_l)}^{+} \sum_{j\in R_k(\tau_l)} (\pi_{ik}^{(1,\,l-1)}-p_k)(\pi_{kj}^{(l)}-\beta_0 p_j)+ \\ +\sum_{k\in R(\tau_l)}^{-} \sum_{j\notin R_k(\tau_l)} (\pi_{ik}^{(1,\,l-1)}-p_k)(\pi_{kj}^{(l)}-\beta_0 p_j),$$

where Σ^+, respectively Σ^-, denote summation over those k for which $\pi_{ik}^{(1,\,l-1)} > p_k$, respectively, $\pi_{ik}^{(1,\,l-1)} < p_k$. By (4.10), the second term in (4.12) does not exceed $\beta_0 e^{-n}$. Using (4.10) again, it is easily proved that the first term in (4.12) does not exceed $(1-\beta_0)\operatorname{Var}(P_i^{(1,l)}, P)+\Delta'/6$. The lemma has been proved.

Applying Lemma 4.3 l times in succession for fixed i, we obtain the estimate

$$(4.13)\qquad \operatorname{Var}(P_i^{(1,\,L)}, P)\leqslant (1-\beta_0)^L+6\sum_{l=1}^{L}\sum_{j\notin R(\tau_l)} \pi_{ij}^{(1,\,l-1)}+6Le^{-n}$$

(here $\pi_{ij}^{(1,\,0)}$ denote the entries of the identity $(I\times I)$-matrix).

The form of the second term on the right-hand side of (4.13) allows us to define the set R_* in Theorem 4.1 as follows. For each $l = 1, 2, \ldots, L$ we choose a subset $R_*(l) \subset \mathcal{I}$ of i's such that

$$(4.14)\qquad \sum_{j\notin R(\tau_l)} \pi_{ij}^{(1,\,l-1)} \leqslant e^{-n/2}$$

and we put $R_* = R_*(1) \cap R_*(2) \cap \ldots \cap R_*(L)$. Now, if $i \in R_*$, then by (4.13) $\operatorname{Var}(P_i^{(1,L)}, P) \leqslant \Delta/2$, that is, the inequality (4.9) (and hence also (4.5)) has been proved. Finally, using (4.14) we obtain

$$\sum_{i\notin R_*} p_i\leqslant e^{n/2}\sum_{l=1}^{L}\sum_{i\notin R_*(l)}\sum_{j\notin R(\tau_l)} p_i\pi_{ij}^{(1,\,l-1)}\leqslant e^{n/2}\sum_{l=1}^{L}\sum_{j\notin R(\tau_l)} p_j\leqslant LNe^{-n/2},$$

which proves (4.6) in the case $k = l+1$.

Theorem 4.1 has been proved completely.

4.2. Construction of the initial lattice.

The Markov lattice is constructed in two steps. In the first step we turn the pre-Markov partition, constructed in [8], into some "lattice construct" which will be a rather rough approximation to the Markov lattice. We call the object arrived at the *initial lattice*. In the second step, in §4.3, we give two sucessive modifications of the initial lattice and obtain as the result the *Markov latttice*. In §4.4 we will separately prove the regularity Property ML4.

We recall the basic properties of a pre-Markov partition for two-dimensional hyperbolic billiards.

Let $m \geqslant 0$ and $m_1, m_2 \geqslant m$ be integers. A *pre-Markov partition* ξ *of the space* M *for the map* T^m is a covering of this space by finitely or countably many curvilinear polygons (with C^1-smooth sides) having the following properties:

1. Distinct polygons may intersect along a boundary only.
2. The sides of the polygons $U \in \xi$ lie on discontinuity curves or on certain LUM and LSM. Correspondingly, the boundary $\partial\xi$ of the partition splits up into three parts, $\partial\xi = \partial^R\xi \cup \partial^u\xi \cup \partial^s\xi$, where $\partial^R\xi$ coincides with R_{-m_1, m_2}, and $\partial^u\xi$ ($\partial^s\xi$) consists of *finitely many* LUM (LSM).
3. $T^m(\partial^s\xi) \subseteq \partial^s\xi$ and $T^{-m}(\partial^u\xi) \subseteq \partial^u\xi$.
4. Any segment of an LUM (LSM) that is part of $\partial^u\xi$ ($\partial^s\xi$) ends either on $\partial^R\xi$ or *strictly inside* some LSM that is part of $\partial^s\xi$ (some LUM that is part of $\partial^u\xi$).

In fact, property 3 of boundary invariance is characteristic for a pre-Markov partition. Once again we stress that the elements of a pre-Markov partition are not parallelograms.

In [8] the partition ξ was constructed for some $m \geqslant 1$ and $m_1 = m_2 = m$, and m could be chosen arbitrarily large. We will assume that m is given. Moreover, for each such m the pre-Markov partition depends only on a small parameter ε, $0 < \varepsilon < \varepsilon_0(m)$. In particular, the diameters of all elements $U \in \xi = \xi(\varepsilon)$ do not exceed $\mathrm{const}(Q)\sqrt{\varepsilon}$.

By property 4, each element $U \in \xi$ not adjacent to $\partial^R\xi$ is a quadrangle. By property 2 there are finitely many such elements. As in [8] we will call them *non-bordering*, and the remaining elements *bordering*. The measure of the union of all bordering elements does not exceed $m\,\mathrm{const}(Q)\varepsilon$.

From the partition $\xi = \xi(\varepsilon)$ we go over to the partition $\xi_1 = \bigvee_{-m}^{m} T^i\xi$. It also has the properties 1–4, for $m_1 = m_2 = 2m$, and is, moreover, pre-Markov for T, and not for T^m. It is easily seen that the measure of the union of the bordering elements of ξ_1 does not exceed $m^2\,\mathrm{const}(Q)\varepsilon$.

A basic step in constructing the initial lattice consists in the fact that in each bordering element $U \in \xi_1$ we draw *the whole* HLUM and *the whole* HLSM *stetched on* U. By intersection we obtain the *parallelogram* $W(U)$. We extract from them the parallelograms satisfying the following three additional conditions:

(A1) the polygon U and its images T^iU for $|i| \leqslant m$ do not intersect subdividing curves;

(A2) $\nu(W(U))/\nu(U) \geqslant 1 - \varepsilon^{b_1}$;

(A3) $p(\gamma^u_{W(U)}(x)) \geqslant \varepsilon^{b_2}$ and $p(\gamma^s_{W(U)}(x)) \geqslant \varepsilon^{b_2}$ for every point $x \in W(U)$.

Here and in the sequel, $b_1, b_2, \ldots$ denote positive constants determined by the choice of the quantities ν and n_0 in §3. The values of these constants will be specified below, while at present we reserve the right to choose them freely.

We define the initial lattice $\mathcal{R}_\varepsilon$ as the set of all parallelograms $W_1, W_2, \ldots, W_I$, $I = I(\varepsilon)$, constructed above and satisfying the conditions (A1)–(A3); we put $W_0 = \overline{M} \setminus \bigcup W_i$. For each $W \in \mathcal{R}_\varepsilon$ we denote by $\tilde{K}(W)$ the element of ξ_1 from which the parellelogram W is obtained.

***Lemma* 4.4.** *The initial lattice $\mathfrak{R}_\varepsilon$ has the following properties:*

(a) (structure) *the parallelograms $W \in \mathfrak{R}_\varepsilon$ are* 0*-homogeneous, and are formed by intersection of all the* HLUM *and all the* HLSM *stretched on $\tilde{K}(W)$; the quadrangles $\tilde{K}(W)$ can intersect only along their boundaries*;

b) (measure of the remainder) $\nu\ (W_0) < \varepsilon^{b_3}$;

c) (Markov property in one step) *for any two parallelograms W', $W'' \in \mathfrak{R}_\varepsilon$ the intersections $TW' \cap W''$ and $T\tilde{K}(W') \cap \tilde{K}(W'')$ either have measure* 0 *or are regular* (*see below for its clarification*);

d) (density) *for each $W \in \mathfrak{R}_\varepsilon$ there are* HLUM $\tilde{\gamma}^u\ (W)$ *and* HLSM $\tilde{\gamma}^s\ (W)$, *stretched on $\tilde{K}(W)$, on which W has density at least $1-\tilde{\delta}_1(\varepsilon)$, where $\tilde{\delta}_1(\varepsilon) \to 0$ as $\varepsilon \to 0$;*

e) (dimensions) *for each $W \in \mathfrak{R}_\varepsilon$ we have* diam $W \leqslant \text{const}(Q)\sqrt{\varepsilon}$, *but $p\ (\tilde{\gamma}^u\ (W)) \geqslant \varepsilon^{b_2}$ and $p\ (\tilde{\gamma}^s\ (W)) > \varepsilon^{b_2}$.*

We recall [8] that if W', W'' are two parallelograms, then the intersection $TW' \cap W''$ is said to be *regular* if it is non-empty and can be represented in the form $[\gamma^u_{W''}\ (x),\ T\gamma^s_{W'}\ (T^{-1}x)]$ for any point $x \in TW' \cap W''$. Similarly, we call the intersection $T\tilde{K}(W') \cap \tilde{K}(W'')$ *regular* if it is a non-empty quadrangle with s-faces on $\partial\tilde{K}(W'')$, while $\tilde{K}(W') \cap T^{-1}\tilde{K}(W'')$ must be a quadrangle with u-faces on $\partial\tilde{K}(W')$. Regularity of intersection characterizes the elements of a Markov partition [8], hence we have called property c) the *Markov property.*

We note that property c) *does not imply* regularity of the intersections $T^nW' \cap W''$ for $n \geqslant 2$, since such an intersection may contain a point x such that $T^{-k}x \in W_0$ for some $1 \leqslant k \leqslant n-1$.

Proof of the lemma. Properties a), c), d), and e) follow directly from the construction. We only note that property d) follows from (A2) and the absolute continuity of addition on LUM and LSM. We also note that the quadrangle $\tilde{K}(W)$ does not always coincide with the carrier of the parallelogram W, since the sides of the element $\tilde{K}(W) \in \xi_1$ can be *inhomogeneous* LUM (LSM).

It remains to prove property b). The set W_0 is the union of all bordering elements $U \in \xi_1$, all non-bordering elements $U \in \xi_1$ not satisfying (A1)$-$(A3), and all complements $U \setminus W(U)$ in those elements satisfying (A1)$-$(A3). An estimate of the total measure of the bordering elements is given below. By (A2), the complements $U \setminus W(U)$ have total measure less than ε^{b_1}.

We consider the elements $U \in \xi_1$ not satisfying (A1). The smallness of the diameters of these elements and the construction of the subdividing curves imply that their total measure does not exceed $2mC_6\varepsilon^{b_4}$, where $C_6 = C_6(\nu, n_0)$ and $b_4 = b_4(\nu) > 0$.

We turn to condition (A2). By construction, the points $x \in U \setminus W(U)$ have short HLUM or HLSM, that is, $\min\{r^{0u}(x),\ r^{0s}(x)\} \leqslant \text{const}\ \sqrt{\varepsilon}$. By Theorem 3.10 the total measure of the elements not satisfying (A2) does not exceed const $(Q)\ \varepsilon^{\theta/2-b_1}$, and it suffices now to put $b_1 = \theta/4$.

Condition (A3) presents the greatest difficulties. It suffices to consider the elements $U \in \xi_1$ that do not satisfy (A3) but do satisfy (A2). We first estimate the measure of each of these. Let $x \in W(U)$ be a point such that $p\,(\gamma^u_{W(U)}\,(x)) \leqslant \varepsilon^{b_2}$ (the other case, when $p\,(\gamma^s_{W(U)}\,(x)) \leqslant \varepsilon^{b_2}$, is completely similar). Using the 0-homogeneity of $W(U)$ we obtain $p\,(\gamma^u_{W(U)}\,(y)) \leqslant C_4^{-1}\varepsilon^{b_2}$ for any point $y \in W(U)$, whence (see (2.3)) $l\,(\gamma^u_{W(U)}\,(y)) \leqslant \text{const}\,(Q)\,\varepsilon^{b_2/2}$ and therefore (see §3.1)

$$\nu\,(W\,(U)) \leqslant \text{const}\,(Q,\,\nu,\,n_0)\,\varepsilon^{b_2/2}. \tag{4.15}$$

Below we will estimate the number N_ε of all non-bordering elements $U \in \xi_1$. Our idea is that their dimensions have order ε, while $\dim M = 2$, hence their number, N_ε, must grow like ε^{-2} as $\varepsilon \to 0$. For our purposes any power-like estimate $N_\varepsilon < \varepsilon^{-b_5}$, $b_5 > 0$, would suffice.

To derive this estimate we have to complete our excursion into the construction of the pre-Markov partition ξ (see [8], §3, §4, §6). We consider a scattering billiard with finite horizon. The construction of ξ starts by choosing two $(C_0\varepsilon)$-nets in certain subsets of M. If these nets are chosen to be minimal, the number of their elements does not exceed $\text{const}(Q)\varepsilon^{-2}$. The requirement of minimality has no influence on the construction of ξ, hence we may impose it here.

For each point of these nets, 7 segments of LUM and LSM of definite shape were constructed, and their images under the action of $T^{\pm m}$ were adjoined to these. For small ε each of the original segments intersects at most mK_0 discontinuity curves in $R_{-m,m}$. Hence the total number of all segments constructed does not exceed $mc'(Q)\varepsilon^{-2}$. Precisely these constitute the boundaries $\partial^u\xi$ and $\partial^s\xi$. The boundary $\partial\xi_1$ consists of just the segments and their images under the action of T^i, $|i| \leqslant m$. By similar arguments, their number does not exceed $m^2c''(Q)\varepsilon^{-2}$. Finally, the total number of non-bordering elements $U \in \xi_1$ does not exceed $m^8c'''(Q)\varepsilon^{-8}$, since each of them is uniquely determined by the set of four segments of the LUM and LSM bounding it.

The case of billiards with infinite horizon (see [8], §4) or with focussing boundary components (see [8], §6) is somewhat more complicated. We leave out a detailed discussion of these cases, it being based on the same arguments. In all cases we obtain the estimate $N_\varepsilon \leqslant \varepsilon^{-b_5}$, $b_5 > 0$.

Together with (4.15), the last estimate shows that the total measure of the elements $U \in \xi_1$ that do not satisfy (A3) but do satisfy (A2) does not exceed $\varepsilon^{b_2/2-b_5}$, and it suffices to put $b_2 = 2(b_5+1)$ and, finally, $b_3 = \min\{b_4,\,b_1,\,1\}$.

Lemma 4.4 has been proved.

4.3. Transition from the initial lattice to the Markov lattice.
Here we give two successive modifications of the initial lattice $\mathscr{R}_\varepsilon$, and obtain as a result the Markov lattice $\mathscr{R}_{n,N}$.

The *first modification* consists in refining the elements $W \in \mathcal{R}_\varepsilon$ to obtain n-homogeneous parallelograms.

We consider all possible intersections

$$(4.16) \qquad T^{-n}W_{i_{-n}} \cap \cdots \cap T^{-1}W_{i_{-1}} \cap W_{i_0} \cap TW_{i_1} \cap \cdots \cap T^n W_{i_n}$$

having non-zero measure (here $W_i \in \mathcal{R}_\varepsilon$). We denote these by $\tilde{V}_i$, $i = 1, 2, \ldots, I(\varepsilon, n)$, and we put $\tilde{V}_0 = \overline{M \setminus \bigcup_i \tilde{V}_i}$. We denote the set $\{\tilde{V}_i\}$, $1 \leqslant i \leqslant I(\varepsilon, n)$, by $\mathcal{R}_{\varepsilon,n}$, and call it a *pre-Markov lattice* (*with parameters* ε, n). For the element $\tilde{V} \in \mathcal{R}_{\varepsilon,n}$ given by (4.16) we put $\tilde{K}(\tilde{V}) = T^{-n}\tilde{K}(W_{i_{-n}}) \cap \ldots \cap T^n\tilde{K}(W_{i_n})$, and also $\tilde{\gamma}^u(\tilde{V}) = T^{-n}\tilde{\gamma}^u(W_{i_{-n}})$ and $\tilde{\gamma}^s(\tilde{V}) = T^n\tilde{\gamma}^s(W_{i_n})$. By Lemma 4.4, the elements $\tilde{V} \in \mathcal{R}_{\varepsilon,n}$ are parallelograms, and $\tilde{K}(\tilde{V})$ is a quadrangle.

***Lemma* 4.5.** *The pre-Markov lattice $\mathcal{R}_{\varepsilon,n}$ has the following properties:*

a) (structure) *the parallelograms $\tilde{V} \in \mathcal{R}_{\varepsilon,n}$ are n-homogeneous, and the quadrangles $\tilde{K}(\tilde{V})$ can intersect only along their faces;*

b) (measure of the remainder) $\nu(\tilde{V}_0) \leqslant 3n\varepsilon^{b_3}$;

c) (Markov property in $\pm n$ steps) *for any two parallelograms $\tilde{V}'$, $\tilde{V}'' \in \mathcal{R}_{\varepsilon,n}$ and any k, $1 \leqslant k \leqslant n$, the intersections $T^k\tilde{V}' \cap \tilde{V}''$ and $T^k\tilde{K}(\tilde{V}') \cap \tilde{K}(\tilde{V}'')$ either have measure 0 or are regular;*

d) (density) *the canonical projection of any parallelogram $\tilde{V} \in \mathcal{R}_{\varepsilon,n}$ onto the* HLUM $\tilde{\gamma}^u(\tilde{V})$ (HLSM $\tilde{\gamma}^s(\tilde{V})$) *has density on this* HLUM (HLSM) *at least* $1 - \tilde{\delta}_2(\varepsilon)$, *where* $\tilde{\delta}_2(\varepsilon) \to 0$ *as* $\varepsilon \to 0$;

e) (dimensions) *for each $\tilde{V} \in \mathcal{R}_{\varepsilon,n}$ we have* $\operatorname{diam} \tilde{V} \leqslant \text{const}\sqrt{\varepsilon}$, *but* $p(T^n\tilde{\gamma}^u(\tilde{V})) \geqslant \varepsilon^{b_2}$ *and* $p(T^{-n}\tilde{\gamma}^s(\tilde{V})) \geqslant \varepsilon^{b_2}$.

Lemma 4.5 follows directly from Lemma 4.4.

The *second modification* pursues its aim by guaranteeing the Markov property in $\pm n$ steps.

For each $\tilde{V}_i \in \mathcal{R}_{\varepsilon,n}$ we define a subparallelogram $V_i = \{x \in \tilde{V}_i : T^k x \notin \tilde{V}_0$ for all $|k| \leqslant N\}$ (that is, we remove from $\tilde{V}_i$ those points which in $\pm N$ steps have at least once "jumped out" of the pre-Markov lattice $\mathcal{R}_{\varepsilon,n}$). In the sequel we consider only those parallelograms $V_i \subset \tilde{V}_i$ for which the inequality

$$(4.17) \qquad \nu(V_i) > (1 - \varepsilon^{b_6})\,\nu(\tilde{V}_i)$$

holds for $b_6 = b_3/2$.

For convenience of notation we enumerate all parallelograms $\tilde{V}_i \in \mathcal{R}_{\varepsilon,n}$ such that (4.17) holds for $i = 1, 2, \ldots, I(\varepsilon, n, N)$ (of course, $I(\varepsilon, n, N) \leqslant I(\varepsilon, n)$). We denote by $\mathcal{R}_{\varepsilon,n,N}$ the set of parallelograms V_i for $1 \leqslant i \leqslant I(\varepsilon, n, N)$, and put $V_0 = \overline{M \setminus \bigcup_i V_i}$.

***Lemma* 4.6.** *The system of sets $\mathcal{R}_{\varepsilon,n,N}$ has the following properties:*

a) (structure) *the parallelograms $V \in \mathcal{R}_{\varepsilon,n,N}$ are n-homogeneous;*

b) (measure of the remainder) $\nu(V_0) \leqslant Nn\varepsilon^{b_6}$;

c) (Markov property in $\pm N$ steps) *for any two parallelograms $V', V'' \in \mathcal{R}_{\varepsilon,n,N}$ and any k, $1 \leqslant k \leqslant N$, the intersection $T^k V' \cap V''$ either has measure 0 or is regular;*

d) (dimensions) *for each $V \in \mathcal{R}_{\varepsilon,n,N}$ we have* $\operatorname{diam} V \leqslant \text{const}\sqrt{\varepsilon}$.

Lemma 4.6 follows directly from Lemma 4.5.

We define the Markov lattice $\mathcal{R}_{n,N}$ as the system $\mathcal{R}_{\varepsilon,n,N}$ for $\varepsilon^{b_7} = e^{-n}$, where $b_7 = \min\{b_2, b_3, b_6\}/2$, and adjoin to it the set V_0.

The Properties ML1 and ML2 follow directly from Lemma 4.6. For proving ML3 it suffices to note that the parallelogram $T^{-i_1-i_{l-1}}V_{j_1} \cap \ldots \cap T^{i_{l-2}-i_{l-1}}V_{j_{l-2}}$ (see (4.1)) is s-inscribed in $V_{j_{l-1}}$, while the parallelogram $T^{i_l-i_{l-1}}V_{j_l} \cap \ldots \cap T^{i_k-i_{l-1}}V_{j_k}$ is u-inscribed in $V_{j_{l-1}}$, and then apply (3.7). The next subsection is devoted to the proof of the regularity Property ML4.

4.4. Proof of the regularity property.

We begin with the following lemma.

***Lemma* 4.7** (on intersections). *For all $\tilde V', \tilde V'' \in \mathcal{R}_{\varepsilon,n}$ and all integers $k \geqslant k_0 = D_1(n-\ln\varepsilon)$ we have $\nu(T^k\tilde V' \cap \tilde V'') \geqslant \beta_1\nu(\tilde V')\nu(\tilde V'')$. Here $\beta_1 > 0$ and $D_1 > 0$ are constants determined by the choice of the quantities ν and n_0 in* §3.

We stress that here the discussion is about the elements of a *pre-Markov* (and not a Markov) lattice. We also note that the case $\tilde V' = \tilde V''$ is not excluded.

Proof. We fix a 0-homogeneous parallelogram U_0 of non-zero measure, and in it a subparallelogram $U_{00} \subset U_0$, also of non-zero measure, whose s- and u-faces lie strictly inside the carrier $K(U_0)$. For each $k \geqslant 1$ we denote by $\tilde\gamma'_{i,k}$, $i \geqslant 1$, all segments inside the HLUM $\tilde\gamma^u(\tilde V')$ whose images under the action of T^k are stretched on U_0 and, moreover, intersect the s-faces of the interior parallelogram U_{00}. By Lemma 4.5e) and Theorem 3.19,

$$p\Big(\bigcup_i \tilde\gamma'_{i,k}\Big)/p(\tilde\gamma(\tilde V')) \geqslant C_4^{-1}\delta_1(U_0) \tag{4.18}$$

for all $k \geqslant k_1 = n - [C^{(1)}b_2 \ln\varepsilon] + n_1(U_0)$.

For each segment $\tilde\gamma'_{i,k}$ we choose a subparallelogram $\tilde V'_{i,k} \subset \tilde V'$ which is canonically projected onto this segment. In the sequel we consider only those i for which the projection of $\tilde V'_{i,k}$ onto $\tilde\gamma'_{i,k}$ has on it density at least $1-\tilde\delta_3$, where $\tilde\delta_3 = \tilde\delta_2^{1/2}$. By (4.18) and Lemma 4.5d), the relative measure of the union of these segments on the curve $\tilde\gamma^u(\tilde V')$ for sufficiently small ε is not less than $\delta_1(U_0)/2$. Hence

$$\nu\Big(\bigcup_i \tilde V'_{i,k}\Big)/\nu(\tilde V') > \delta_2(U_0) \tag{4.19}$$

for some $\delta_2(U_0) > 0$, provided that ε is sufficiently small.

Moreover, by Lemma 3.7 the density of the canonical projection onto the HLUM $T^k\widetilde{\gamma}'_{i,k}$, stretched on U_0, of the parallelograms $T^k\tilde{V}'_{i,l}$ is at least $1-C_4^{-1}\tilde{\delta}_3$. Hence these parallelograms are sufficiently "elongated", that is, their s-faces almost reach the s-faces of U_0, and at the same time their u-faces remain substantially away from the u-faces of the latter, since the curves $T^k\widetilde{\gamma}'_{i,k}$ intersect the s-faces of U_{00}. We call this property *regular position.*

We consider a parallelogram $\tilde{V}''$ under the conditions of Lemma 4.7. In a completely similar way we can construct parallelograms $\tilde{V}''_{j,l} \subset \tilde{V}''$ for $l \leqslant -k_1$ (that is, for negative iterations of T).

We study the intersection of the parallelograms $U'_i = T^k\tilde{V}'_{i,k}$ and $U''_j = T^l\tilde{V}''_{j,l}$. By the property of "regular position", the quadrangles $K(U'_i)$ and $K(U''_j)$ intersect in a regular manner, that is, their intersection is bounded by the u-faces of $K(U'_i)$ and the s-faces of $K(U''_j)$. This and Lemmas 4.4 and 4.5 imply that the intersection $U'_i \cap U''_j$ of the parallelograms is also regular, that is, it is non-empty and can be represented as $[\gamma^u_{U''_j}(x), \gamma^s_{U'_i}(x)]$ for an arbitrary point $x \in U'_i \cap U''_j$. We note that this assertion would be not true if we used elements of a *Markov* lattice instead of those of a *pre-Markov* lattice, since then the intersection $U'_i \cap U''_j$ could be empty.

Remark 4.8. For each pair i, j the parallelogram $W_{ij} = T^{-k}(U'_i \cap U''_j)$ is s-inscribed in $\tilde{V}'$, and the parallelogram $T^{-l}(U'_i \cap U''_j)$ is u-inscribed in $\tilde{V}''$, and, moreover, $T^{k-l}W_{ij} \subset \tilde{V}''$. In particular, for $\tilde{V}' = \tilde{V}''$ the intersection $T^{k-l}W_{ij} \cap W_{ij}$ is non-empty and regular. We will use this property in §7.

Since U_0 is 0-homogeneous, we have $\nu(U'_i \cap U''_j) \geqslant \delta_3\nu(U'_i)\nu(U''_j)$, where $\delta_3 = \delta_3(U_0) > 0$. Summation over all i, j and subsequent application of (4.19) (and the similar estimate for parallelograms $\tilde{V}''_{j,l}$) leads to $\nu(T^k\tilde{V}' \cap T^l\tilde{V}'') \geqslant \beta_1\nu(\tilde{V}')\nu(\tilde{V}'')$ for $\beta_1 = \delta_3\delta_2^2$. Since $k \geqslant k_1$ and $l \leqslant -k_1$ are arbitrary, we obtain Lemma 4.7 with $k_0 = 2k_1$.

It remains to derive Property ML4. It follows from Lemmas 4.7 and 4.5b) and relation (4.17) if we put $\beta_0 = \beta_1/2$ and $D_0 = 3(1+2C^{(1)}b_2/b_7)$.

§5. An estimate for the decay of correlations

Here we will derive Theorem 1.1, using the constructions of §3–§4. We choose a sufficiently large N, put $n = [\sqrt{N}]$, and consider the Markov lattice $\mathcal{R}_{n,N}$. Averaging the function $F(x)$ from the conditions of Theorem 1.1 over the elements of $\mathcal{R}_{n,N}$, we obtain a step-function $\tilde{F}(x)$, generating a new stationary process $\tilde{X}_k = \tilde{F}(T^kx)$.

Using Properties ML1 and ML2 and the fact that F is Hölder continuous, we can easily prove the estimate

$$(5.1) \qquad |\langle X_0 \cdot X_N\rangle - \langle \tilde{X}_0 \cdot \tilde{X}_N\rangle| \leqslant C_1(F)\,\gamma_1^n$$

for some $\gamma_1(\alpha) < 1$ depending only on the index α in the Hölder condition; moreover, $\gamma_1(\alpha) \to 1$ as $\alpha \to 0$ (here we use the fact that, since $\overline{M}$ is compact, $\sup|F(x)| < \infty$).

We denote by f_i the value of $\tilde{F}(x)$ at the element $V_i \in \mathfrak{R}_{n,N}$. We have

$$\langle \tilde{X}_0 \cdot \tilde{X}_N \rangle = \sum_{i,j=0}^{I} f_i f_j \nu(V_i \cap T^N V_j) = \sum_{i,j=0}^{I} f_i f_j \nu(V_i / T^N V_j)\, \nu(V_j).$$

Applying Theorem 4.1 and using the fact that $N \geqslant n^2$, we obtain

$$\langle \tilde{X}_0 \cdot \tilde{X}_N \rangle = \sum_{i,j=0}^{I} f_i f_j \nu(V_i)\, \nu(V_j) + O(\gamma_2^n) \tag{5.2}$$

for some $\gamma_2 < 1$ that does not depend on F. The first term on the right-hand side of (5.2) vanishes, since $\langle F(x) \rangle = \langle \tilde{F}(x) \rangle = 0$. The estimates (5.1) and (5.2) lead to Theorem 1.1. The restriction on the Hölder index α for F must be chosen such that $\gamma_1(\alpha) \leqslant \gamma_2$.

The proof of the generalization of Theorem 1.1 to two *distinct* functions F and G (inequality (1.4)) goes through unchanged.

Remark 5.1. The requirements on the function F (and G in (1.4)) can easily be weakened: assume it to be "piecewise Hölder", that is, to satisfy the Hölder condition (with index α) on finitely many subdomains $M_1, \ldots, M_l$ in M with piecewise smooth boundaries and such that $\overline{M} = \bigcup \overline{M}_i$. In this case the proof of Theorem 1.1 requires only one addition in the derivation of (5.1): the elements of the Markov lattice intersecting $\bigcup \partial M_i$ have total measure $\leqslant C(F)e^{-n}$.

In the case of scattering billiards with finite horizon, this remark allows us to estimate the decay of correlations for functions of the form $F(x) = \Phi(T^{-m}x, \ldots, x, Tx, \ldots, T^m x)$, where Φ is a smooth function of several variables. Such functions can have singularities on $R_{-m,m}$ only. The function $\tau(x)$ is an example of such a function.

§6. The central limit theorem

Our proof of Theorem 1.2 uses Bernstein's classical method [3], as do the similar proofs in [9], [23].

We must immediately note that by Theorem 1.1 the quantity σ^2 in (1.2) is finite, and, moreover, if $\sigma \neq 0$, then $\mathrm{D}S_N = \langle S_N^2 \rangle = \sigma^2 N(1 + o(1))$. Therefore (1.3) is equivalent to

$$\nu\left\{\frac{1}{\sqrt{\mathrm{D}S_N}} S_N < z\right\} \xrightarrow[N\to\infty]{} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-u^2/2}\, du$$

(in fact, it is precisely this relation that is called the *central limit theorem*, see [9]).

6.1 An estimate of the fourth moment.
We begin with the following lemma.

***Lemma* 6.1.** $\langle S_N^4 \rangle \leqslant C(F)N^2$.

For the proof we write

$$\langle S_N^4 \rangle = \sum_{t_1, t_2, t_3, t_4=0}^{N} \langle X_{t_1} X_{t_2} X_{t_3} X_{t_4} \rangle \tag{6.1}$$

and estimate the quantity $\langle X_{t_1} X_{t_2} X_{t_3} X_{t_4} \rangle$, assuming that $t_1 \leqslant t_2 \leqslant t_3 \leqslant t_4$.

First we single out those terms in (6.1) for which $|t_4 - t_1| \leqslant 3N^{1/3}$. Their number does not exceed $10N^2$, hence it suffices to consider only the remaining terms. These split into three groups:

a) $|t_2 - t_1| \geqslant N^{1/3}$. We put $n = [\sqrt{t_2 - t_1}]$ and consider the Markov lattice $\mathcal{R}_{n,N}$. As in §5, we replace $F(x)$ by $\tilde{F}(x)$ (and thus X_i by $\tilde{X}_i$); here

$$|\langle X_{t_1} X_{t_2} X_{t_3} X_{t_4} \rangle - \langle \tilde{X}_{t_1} \tilde{X}_{t_2} \tilde{X}_{t_3} \tilde{X}_{t_4} \rangle| \leqslant C(F)Ne^{-n}.$$

We have

$$\langle \tilde{X}_{t_1} \tilde{X}_{t_2} \tilde{X}_{t_3} \tilde{X}_{t_4} \rangle = \sum_{i_1, i_2, i_3, i_4=0}^{I} f_{i_1} f_{i_2} f_{i_3} f_{i_4} \nu(T^{-t_1} V_{i_1} \cap T^{-t_2} V_{i_2} \cap T^{-t_3} V_{i_3} \cap T^{-t_4} V_{i_4}) =$$
$$= \sum f_{i_1} f_{i_2} f_{i_3} f_{i_4} \nu(T^{-t_1} V_{i_1} / T^{-t_2} V_{i_2} \cap T^{-t_3} V_{i_3} \cap T^{-t_4} V_{i_4}) \nu(T^{-t_2} V_{i_2} \cap T^{-t_3} V_{i_3} \cap T^{-t_4} V_{i_4})$$

Using Theorem 4.1 (see also Remark 4.2) and the fact that $|t_2 - t_1| \geqslant n^2$, we obtain

$$\langle \tilde{X}_{t_1} \tilde{X}_{t_2} \tilde{X}_{t_3} \tilde{X}_{t_4} \rangle = \sum f_{i_1} f_{i_2} f_{i_3} f_{i_4} \nu(T^{-t_1} V_{i_1}) \times \tag{6.2}$$
$$\times \nu(T^{-t_2} V_{i_2} \cap T^{-t_3} V_{i_3} \cap T^{-t_4} V_{i_4}) + O(\gamma^n)$$

for some $\gamma \in (0, 1)$ that does not depend on F. But the first term on the right-hand side of (6.2) vanishes, since $\langle F(x) \rangle = \langle \tilde{F}(x) \rangle = 0$; hence

$$|\langle \tilde{X}_{t_1} \tilde{X}_{t_2} \tilde{X}_{t_3} \tilde{X}_{t_4} \rangle| \leqslant C(F) \gamma^{\sqrt{t_2 - t_1}}.$$

b) $|t_3 - t_2| \geqslant N^{1/3}$. We put $n = [\sqrt{t_3 - t_2}]$ and consider the Markov lattice $\mathcal{R}_{n,N}$. Acting as in the previous group, we write

$$\langle \tilde{X}_{t_1} \tilde{X}_{t_2} \tilde{X}_{t_3} \tilde{X}_{t_4} \rangle = \sum_{i_1, i_2, i_3, i_4=0}^{I} f_{i_1} f_{i_2} f_{i_3} f_{i_4} \nu(T^{-t_1} V_{i_1} \cap T^{-t_2} V_{i_2} / T^{-t_3} V_{i_3} \cap T^{-t_4} V_{i_4}) \times$$
$$\times \nu(T^{-t_3} V_{i_3} \cap T^{-t_4} V_{i_4}).$$

Using Theorem 4.1 and the fact that $|t_3 - t_2| \geqslant n^2$, we obtain

$$\langle \tilde{X}_{t_1} \tilde{X}_{t_2} \tilde{X}_{t_3} \tilde{X}_{t_4} \rangle = \sum_{i_1, i_2, i_3, i_4=0}^{I} f_{i_1} f_{i_2} f_{i_3} f_{i_4} \nu(T^{-t_1} V_{i_1} \cap T^{-t_2} V_{i_2}) \nu(T^{-t_3} V_{i_3} \cap T^{-t_4} V_{i_4}) + O(\gamma^n)$$

for some $\gamma \in (0, 1)$. This equality can be written as

$$\langle \tilde{X}_{t_1} \tilde{X}_{t_2} \tilde{X}_{t_3} \tilde{X}_{t_4} \rangle = \langle \tilde{X}_{t_1} \tilde{X}_{t_2} \rangle \langle \tilde{X}_{t_3} \tilde{X}_{t_4} \rangle + O(\gamma^n),$$

and by Theorem 1.1,

$$|\langle \bar{X}_{t_1}\bar{X}_{t_2}\bar{X}_{t_3}\bar{X}_{t_4}\rangle| \leqslant C(F)\, e^{-\sqrt{t_2-t_1}-\sqrt{t_4-t_3}} + O(\gamma^n).$$

c) $|t_4-t_{43}| \geqslant N^{1/3}$. This can be treated similarly to case a).

Combining these estimates we arrive at the inequality in Lemma 6.1.

Remark. These estimates can be generalized to moments of arbitrary orders. We consider the quantity $\langle X_{t_1}\cdot X_{t_2}\cdot \ldots \cdot X_{t_k}\rangle$ for $t_1 \leqslant t_2 \leqslant \ldots \leqslant t_k$ and order the differences $t_2-t_1, t_3-t_2, \ldots, t_k-t_{k-1}$. From these $k-1$ differences we extract the $[(k+1)/2]$-th (in order of increase), and denote it by d. Then $|\langle X_{t_1}X_{t_2}\ldots X_{t_k}\rangle| \leqslant C(F,k)\,\gamma^d$ for some $\gamma < 1$ that does not depend on F and k. This implies that $|\langle S_N^k\rangle| \leqslant C(F,k)\,N^{[k/2]}$ for any integer $k \geqslant 1$.

6.2. Approximation of the characteristic function.

The main instrument in the proof of limit theorems (going back to S.N. Bernstein) consists in partitioning the whole time interval $\Delta = [1, N]$ into subintervals

$$\Delta = \Delta_1' \cup \Delta_1'' \cup \Delta_2' \cup \Delta_2'' \cup \ldots \cup \Delta_k' \cup \Delta_k'' \cup \Delta_0,$$

where the lengths of the intervals Δ_i' are equal to $L = [N^a]$, the lengths of the intervals Δ_i'' are equal to $l = [N^b]$, and the length of the "remainder" Δ_0 does not exceed $L+l$. We choose a and b such that $1 > a > b > 0$. Then the intervals Δ_i' of greatest length will be separated from each other by the shorter Δ_i''. We put $\Delta' = \Delta_1' \cup \Delta_2' \cup \ldots \cup \Delta_k'$ and $\Delta'' = \Delta_1'' \cup \Delta_2'' \cup \ldots \cup \Delta_k'' \cup \Delta_0$. We have the expansion

$$S_N = \sum_{t\in\Delta} X_t = \sum_{t\in\Delta'} X_t + \sum_{t\in\Delta''} X_t = S_N' + S_N''.$$

The number of terms in S_N'' does not exceed $2N^h$, where

$$h = \max\{a, 1-a+b\} < 1.$$

Hence, by Theorem 1.1, $\mathsf{D}S_N'' = \langle (S_N'')^2\rangle \leqslant C(F)\,N^h$. Using Chebyshev's inequality, we obtain

$$\nu\left\{x\colon \frac{1}{\sigma\sqrt{N}} S_N'' > \varepsilon\right\} \leqslant \langle (S_N'')^2\rangle/(\varepsilon^2\sigma^2 N) \to 0$$

as $N \to \infty$, that is, the quantity $S_N''/\sqrt{\mathsf{D}S_N}$ tends to zero in probability, and, moreover, $\mathsf{D}S_N \sim \mathsf{D}S_N'$. Therefore (see, for example, [9], Lemma 18.4.1) the limiting distributions of the quantities $S_N/\sqrt{\mathsf{D}S_N}$ and $S_N'/\sqrt{\mathsf{D}S_N'}$ coincide. The last quantity can be written in the form

$$\frac{S_N'}{\sqrt{\mathsf{D}S_N'}} = \frac{1}{\sqrt{\mathsf{D}S_N'}}\Big(\sum_{r=1}^{k}\sum_{t\in\Delta_r'} X_t\Big) = \frac{1}{\sqrt{\mathsf{D}S_N'}}\sum_{r=1}^{k} S_L^{(r)}. \tag{6.3}$$

The lengths of the "separating" intervals Δ_r'', $1 \leqslant r \leqslant k$, although relatively small, do tend to infinity as $N \to \infty$, therefore in (6.3) the quantities

$S_L^{(r)}$ for distinct r must be weakly dependent. This allows us to approximate the characteristic function (6.3), which is equal to

$$\varphi(\lambda) = \left\langle \exp\left(\frac{i\lambda}{\sqrt{\mathsf{D}S'_N}} \sum_{r=1}^{k} S_L^{(r)}\right)\right\rangle,$$

by the simpler expression

$$\varphi_0(\lambda) = \left[\left\langle \exp\left(\frac{i\lambda}{\sqrt{k\mathsf{D}S_L^{(1)}}} S_L^{(1)}\right)\right\rangle\right]^k.$$

To this end we introduce the Markov lattice $\mathcal{R}_{n,N}$ with $n = [N^{b/2}]$. We replace $F(x)$ by its average $\tilde{F}(X)$ over the elements of $\mathcal{R}_{n,N}$ and, correspondingly, the process $\{X_t\}$ by $\tilde{X}_t = \tilde{F}(T^t x)$. The quantities $S_L^{(r)}$ in (6.3) are correspondingly replaced by $\tilde{S}_L^{(r)}$. We consider the functions

$$\tilde{\varphi}(\lambda) = \left\langle \exp\left(\frac{i\lambda}{\sqrt{\mathsf{D}S'_N}} \sum_{r=1}^{k} \tilde{S}_L^{(r)}\right)\right\rangle$$

and

$$\tilde{\varphi}_0(\lambda) = \left[\left\langle \exp \frac{i\lambda}{\sqrt{k\mathsf{D}S_L^{(1)}}} \tilde{S}_L^{(1)}\right\rangle\right]^k.$$

Using Properties ML1 and ML2 of a Markov lattice, we can easily prove that $\varphi(\lambda) = \tilde{\varphi}(\lambda) + o(1)$ and $\varphi_0(\lambda) = \tilde{\varphi}_0(\lambda) + o(1)$ as $N \to \infty$.

The function $\tilde{\varphi}(\lambda)$ is piecewise constant, and can be written in the form

$$\tilde{\varphi}(\lambda) = \sum_{r=1}^{k} \sum_{t\in\Delta'_r} \sum_{\alpha_t=0}^{I} \exp\left(\frac{i\lambda}{\sqrt{\mathsf{D}S'_N}} \sum_{r=1}^{k} \sum_{t\in\Delta'_r} f_{\alpha_t}\right) \nu\left(\bigcap_{r=1}^{k} \bigcap_{t\in\Delta'_r} T^{-t} V_{\alpha_t}\right),$$

where, as in §5, the value of $\tilde{F}(X)$ at an element $V_\alpha \in \mathcal{R}_{n,N}$ is denoted by f_α, and $I = I(n, N)$. We replace the unconditional probabilities figuring in this formula by the conditional probabilities

$$\nu\left(\bigcap_{t\in\Delta_1} T^{-t} V_{\alpha_t} / \bigcap_{r=2}^{k} \bigcap_{t\in\Delta'_r} T^{-t} V_{\alpha_t}\right) \nu\left(\bigcap_{r=2}^{k} \bigcap_{t\in\Delta'_r} T^{-t} V_{\alpha_t}\right).$$

Using Theorem 4.1 and the fact that $|\Delta''_r| = l \geqslant n^2$, we obtain

$$\tilde{\varphi}(\lambda) = \sum_{r=1}^{k} \sum_{t\in\Delta'_r} \sum_{\alpha_t=0}^{I} \exp\left(\frac{i\lambda}{\sqrt{\mathsf{D}S'_N}} \sum_{r=1}^{k} \sum_{t\in\Delta'_r} f_{\alpha_t}\right) \nu\left(\bigcap_{t\in\Delta'_1} T^{-t} V_{\alpha_t}\right) \times$$
$$\times\, \nu\left(\bigcap_{r=2}^{k} \bigcap_{t\in\Delta'_r} T^{-t} V_{\alpha_t}\right) + O(\gamma_1^n)$$

for some $\gamma_1 \in (0, 1)$ that does not depend on F. By further continuing this "separation" of the intervals Δ'_r and subsequently using the fact that by Theorem 1.1 $\mathsf{D}S'_N \sim k\mathsf{D}S_L^{(1)}$, we arrive at the estimate $\tilde{\varphi}(\lambda) = \tilde{\varphi}_0(\lambda) + o(1)$ as $N \to \infty$. Hence $\varphi(\lambda) = \varphi_0(\lambda) + o(1)$.

The function $\varphi_0(\lambda)$ is the characteristic function of the sum of k independent random variables, each of which has the same distribution as $\xi_L = S_L^{(1)}/\sqrt{k\mathrm{D}S_L^{(1)}}$. We consider the following rectangular array of independent variables, which are identically distributed in each row:

$$\begin{array}{l} \xi_{L_1}^{(1)}, \ldots, \xi_{L_1}^{(k_1)}, \\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot \\ \xi_{L_N}^{(1)}, \ldots, \xi_{L_N}^{(k_N)}, \\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot \end{array}$$

where $L_N = L$ and $k_N = k$ are defined above, and $\xi_{L_N}^{(i)}$ has the same distribution as ξ_L. It remains to verify that the sums $\xi_L^{(1)} + \ldots + \xi_L^{(k)}$ converge in distribution to the standard normal law. For this it is necessary and sufficient that the following analogue of the Lindeberg condition (see, for example, [9]) holds:

$$k \int\limits_{|z|>\varepsilon} z^2 \, d\nu\, \{\xi_L < z\} \xrightarrow[N\to\infty]{} 0 \tag{6.4}$$

for any $\varepsilon > 0$.

By Lemma 6.1, $\langle S_L^4 \rangle \leqslant C(F)L^2$, hence $\langle \xi_L^4 \rangle \leqslant C(F)L^2/(k\mathrm{D}S_L^{(1)})^2 \leqslant C'(F)/k^2$. This and Chebyshev's inequality imply that $\nu\{x : |\xi_L| > |z|\} \leqslant C'(F)/(kz)^2$, hence

$$k \int\limits_{|z|>\varepsilon} z^2 \, d\nu\, \{\xi_L < z\} \leqslant \frac{C''(F)}{\varepsilon^2 k}$$

and (6.4) holds.

Theorem 1.2 has been proved.

Remark 6.2. The proof of Theorem 6.1 can be carried over unchanged to the case of "piecewise Hölder" functions, which were described in §5 (see Remark 5.1).

§7. Applications. Diffusion in deterministic systems

Here we derive Theorems 1.5 and 1.6 as direct consequences of Theorem 1.2. Regrettably, the main difficulty lies in the verification of the non-degeneracy of the corresponding limiting Gaussian distributions (the condition $\sigma \neq 0$ in Theorem 1.2).

7.1. The distribution of the number of reflections.
We recall that Theorem 1.5 and 1.6 relate to scattering billiards with finite horizon. In this case Theorem 1.2 is applicable to the function $\tau(x) - \langle\tau\rangle$ (see Remark 6.2). As a result we obtain

$$\nu\left\{\frac{\tau(x) + \ldots + \tau(T^{n-1}x) - n\langle\tau\rangle}{\sigma_\tau \sqrt{n}} \leqslant z\right\} \xrightarrow[n\to\infty]{} \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{z} e^{-\frac{u^2}{2}} du, \tag{7.1}$$

where σ_τ is defined by the general formula (1.2).

For proving Theorem 1.5 we write $\nu\{N_t(x) \leqslant z\} = \nu\{\tau(x) + \ldots + \tau(T^n x) > t\}$ for $n = [z]$. This and (7.1) easily imply (1.5) for $a_1 = 1/\langle\tau\rangle$ and $b_1 = \sigma_\tau^2/\langle\tau\rangle^3$.

It remains to prove that $\sigma_\tau \neq 0$. If this were not true, then by Remark 1.4 the function $\tau(x)$ would be homologous to a constant:

$$\tau(x) = \langle\tau\rangle + G(Tx) - G(x) \tag{7.2}$$

for some $G \in L_2(M, \nu)$. Using general facts from ergodic theory, we will show that (7.2) is impossible.

The billiard flow $\{S^t\}$ can be represented as a *special flow*, constructed on the cross-section M from the automorphism t, and with "ceiling" function $\tau(x)$. Relation (7.2) allows us to redefine the cross-section (putting $M_0 = \{S^{-G(x)}x : x \in M\}$) in such a way that the new representation of $\{S^t\}$ as a special flow has constant "ceiling" ($\equiv \langle\tau\rangle$). It is easily seen that a special flow with constant ceiling is not a K-flow, and is not even mixing. However, our flow $\{S^t\}$ has the K-property (see [13]). A contradiction.

7.2. Diffusion in the periodic Lorentz gas.

Theorem 6.1 is just a somewhat strengthened version of Theorem 2 in [23]. In contrast to [23], here we allow break points of the boundary ∂Q and we have weakened condition B (see §2).

The proof of Theorem 6.1 does not differ from that in [23], hence we only give its outline. First we go over from continuous time t to discrete time n, the counter of the number of reflections. Then (1.6) reduces to

$$\nu\left\{\left(\frac{q_1^{(n)}(x)}{\sqrt{n}}, \frac{q_2^{(n)}(x)}{\sqrt{n}}\right) \in A\right\} \xrightarrow[n\to\infty]{} \int_A g_1(q_1, q_2)\, dq_1\, dq_2 \tag{7.3}$$

for some (other) non-degenerate Gaussian distribution g_1 with zero mean; here $(q_1^{(n)}, q_2^{(n)})$ denotes the position of the wandering point at the moment of n-th reflection. In reducing (1.6) to (7.3) we have used Theorem 1.5.

For proving (7.3) we introduce a two-dimensional function $(\Delta_1(x), \Delta_2(x))$ on the space M of the billiard with original domain Q (on the torus). In fact, we put $\Delta_i(x) = \hat{q}_i(x) - q(x)$ for $i = 1, 2$, where $q_1(x)$, $q_2(x)$ are the coordinates of the point x in Q and $\hat{q}_1(x)$, $\hat{q}_2(x)$ are the coordinates of the point of first reflection in ∂Q_∞ of the *lift* of the trajectory of x to the universal covering torus. In other words, $\Delta(x) = (\Delta_1(x), \Delta_2(x))$ is the vector of displacement after one reflection. For $i = 1, 2$ we have

$$q_i^{(n)}(x) = q_i(x) + \Delta_i(x) + \Delta_i(Tx) + \ldots + \Delta_i(T^{n-1}x). \tag{7.4}$$

The first term in (7.4) is uniformly bounded: $|q_i(x)| \leqslant 1$, and Theorem 1.2 is applicable to the remaining sum, since Δ_i is a "piecewise Hölder" function (see Remarks 5.1 and 6.2) with zero mean (this follows from the relation $\Delta_i(x) = -\Delta_i(Tx)$). Moreover, the two-dimensional analogue of Theorem 1.2 (see [23]) is applicable to the two-dimensional function $\Delta(x) = (\Delta_1(x), \Delta_2(x))$;

it gives (7.3) also, provided that the covariance matrix $V = \|v_{ij}\|$ with entries

$$v_{ij} = \sum_{n=-\infty}^{\infty} \langle \Delta_i (x) \cdot \Delta_j (T^n x) \rangle$$

for $1 \leqslant i, j \leqslant 2$ is non-singular.

We show that in fact V is non-singular. Suppose this is not so. Then Remark 1.4 implies that some linear combination of the components $\Delta_1(x)$ and $\Delta_2(x)$ is homologous to zero: $L(x) = a_1\Delta_1(x)+a_2\Delta_2(x) = G(Tx)-G(x)$ for some $G \in L_2(M, \nu)$. In this case (7.4) implies that

$$a_1 q_1^{(n)} (x) + a_2 q_2^{(n)} (x) = G_0 (x) + G (T^n x) - G (x),$$

where $|G_0(x)| \leqslant$ const. This means that the trajectory of a typical point $x \in M$ wanders in Q^∞, remaining "basically" inside some strip in the plane. More precisely, for any $\varepsilon > 0$ there is an A_ε such that $\nu\{|a_1q_1^{(n)}(x)+a_2q_2^{(n)}(x)| \geqslant A_\varepsilon\} < \varepsilon$ for all $n \geqslant 1$. If a_1/a_2 is a rational number, then the projection of the given strip onto a sufficiently large torus with fundamental domain $K_{M,N} = \{0 \leqslant x \leqslant M, 0 \leqslant y \leqslant N\}$, where M, N are integers, forms an A_ε-neighbourhood of a periodic winding of this torus, and covers on it a domain of relatively small area. But this contradicts the ergodicity of the scattering billiard on this torus with reflectors $\partial Q_\infty \cap K_{M,N}$ (see also [23]).

If a_1/a_2 is irrational, then the projection of our strip on any arbitrarily large torus $K_{M,N}$ (with integers M, N) fills it almost densely, and the previous reasoning is no longer sufficient. This case was not sketched in [23], hence we will dwell on it in more detail. Our constructions go back to [11], but are more cumbersome, since we have to deal with discontinuous systems and non-homogeneous hyperbolicity.

***Lemma* 7.1.** *There is a periodic point $y_0 \in M$ such that the sum $S_0 = L(y_0)+ \ldots +L(T^{k-1}y_0)$ is non-zero. Here, k is the period of the point y_0 ($T^k y_0 = y_0$).*

It is easily seen that for any periodic point y_0 the sum S_0 is equal to $a_1m_1(y_0)+a_2m_2(y_0)$ for certain integers $m_1(y_0)$, $m_2(y_0)$. In our case a_1/a_2 is irrational, hence it suffices to find a periodic point y_0 for which m_1 or m_2 is non-zero.

We consider the scattering billiard on the torus with fundamental domain K_{12} (see above) and with reflectors $\partial Q_\infty \cap K_{12}$. In this billiard we denote by Q_{12}, M_{12}, and T_{12} the objects which are denoted by Q, M, and T, respectively, in the original billiard. If $\mathcal{R}_{\varepsilon,n}$ is the pre-Markov lattice of the original billiard in Q, then we can shift it unchanged to the domain $\{0 \leqslant x \leqslant 1, 1 \leqslant y \leqslant 2\}$, and obtain a pre-Markov lattice in M_{12}. Taking an arbitrary parallelogram $V \in \mathcal{R}_{\varepsilon,n}$ and its image V' under this shift, and applying Remark 4.8 to it, we find a parallelogram $V_1 \subset V$ such that $T_{12}^k V_1 \subset V'$ and such that for some k the intersection $T^k V_1 \cap V_1$ is regular.

Then the point $y_0 = \ldots \cap T^{-k}V_1 \cap V_1 \cap T^kV_1 \cap T^{2k}V_1 \cap \ldots$ is periodic, with period k, and $m_2(y_0)$ is odd, that is, non-zero. The lemma has been proved.

***Lemma* 7.2.** *Let y_0 be a periodic point. For any $\varepsilon_0 > 0$ there is a 0-homogeneous parallelogram $U_0 \ni y_0$ such that $y_0 \notin \partial K(U_0)$ (that is, y_0 lies strictly inside $K(U_0)$) and $\nu(U_0)/\nu(K(U_0)) > 1-\varepsilon_0$ (that is, U_0 is "sufficiently dense").*

The proof of this lemma is based on the same arguments as the similar construction in §3.5, and we omit it.

We fix a point y_0 as in Lemma 7.1 and a parallelogram as in Lemma 7.2, for some small ε_0. If U_0 is sufficiently small, then T^kU_0 is also 0-homogeneous and the intersection $T^kU_0 \cap U_0$ is regular.

Now let $G(x) \in L_2(M, \nu)$ be homologous to the function $L(x)$, that is, $L(x) \equiv G(Tx)-G(x)$ almost everywhere in M. If $G(x)$ were essentially bounded, then, as in [11], we would immediately obtain a contradiction to Lemma 7.1. Using only the measurability and integrability of G, we can obtain the following:

***Lemma* 7.3.** *For any $\varepsilon_1 > 0$ there is a 0-homogeneous parallelogram U_1 such that $\nu(U_1)/\nu(K(U_1)) > 1-\varepsilon_1$ (that is, U_1 is "sufficiently dense") and $\nu\{x \in U_1 : |G(x)-g| > \varepsilon_1\}/\nu(U_1) < \varepsilon_1$ for some $g \in \mathbb{R}$ (that is, $G(x)$ can be well approximated on U_1 by the constant function $\tilde{G}(x) \equiv g$).*

We define the "quasiderived" map $T_1 : U_1 \to U_1$ as follows. For a point $x \in U_1$ we find on its trajectory the first image (for $n > 0$) in U_0, and then the first image in U_1. This will be T_1x. The map T_1 is not invertible on the set of points $x \in U_1$ whose trajectories manage to return to U_1 before the first hit on U_0. By making U_1 smaller we can always assume that T_1 is invertible and preserves the measure ν on U_1.

We define two maps, Φ and Φ_1, on U_1. Let $x \in U_1$, let $T^{n_0}x \in U_0$ be the first point of its trajectory in U_0, and let $T^{n_1}x = T_1x$ $(n_1 > n_0)$. We write $\tilde{x} = \gamma^u(T^{n_0+k}x) \cap \gamma^s(T^{n_0}x)$, and set $\Phi x = T^{-n_0-k}\tilde{x}$ and $\Phi_1 x = T^{n_1-n_0}\tilde{x}$. Then Φx is a point which lies, with x, on a single HLUM, and $\Phi_1 x$ lies, with T_1x, on a single HLSM. Moreover, $T^{n_1+k}(\Phi x) = \Phi_1 x$.

We put

$$S(x) = L(x) + \ldots + L(T^{n_1-1}x)$$

and

$$S'(\Phi x) = L(\Phi x) + L(T(\Phi x)) + \ldots + L(T^{n_1+k-1}(\Phi x)).$$

***Lemma* 7.4.** $|S'(\Phi x)-S(x)| \geqslant S_0-\varepsilon_2$, *where ε_2 is sufficiently small (more precisely, $\varepsilon_2 \to 0$ as* diam $U_0 \to 0$).

In fact, the trajectories of x and Φx are close during the first n_0 iterations. Further, the trajectories of $T^{n_0}x$ and $T^{n_0+k}(\Phi x) = \tilde{x}$ are close during n_1-n iterations. The Hölder continuity of $L(x)$ on the continuity set of T implies that the corresponding sums, occurring in the sums $S(x)$ and $S'(\Phi x)$ of interest to us, are close. There remain exactly k terms in $S'(\Phi x)$ not accounted for, coming from the images of $T^{n_0}(\Phi x) = T^{-k}\tilde{x}$. But these images approximate the periodic orbit of y_0, hence the corresponding sum of values of L is close to S_0. The lemma has been proved.

Lemma 7.4 implies the inequality

$$(7.5) \qquad |G(T_1 x) - G(x)| + |G(\Phi_1 x) - G(\Phi x)| \geqslant S_0 - \varepsilon_2.$$

It remains to prove that for the majority (with respect to the measure ν) of points $x \in U_1$ their images Φx and $\Phi_1 x$ coincide in U_1, and that the sets of these images, $\{\Phi x\}$ and $\{\Phi_1 x\}$, have relatively large measure ($>$ const $\nu(U_1)$). Then for such points $\Phi_1 x = T_1(\Phi x)$, and together with (7.5) this leads to a contradiction with the second estimate in Lemma 7.3.

We study the sets ΦU_1 and $\Phi_1 U_1$. We use the notations introduced in the definition of Φ and Φ_1. Since $T^{n_0}x \in U_0$, we find that $\gamma_0 = \gamma^u_{K(U_0)}(T^{n_0}x)$ is an HLUM stretched on $K(U_0)$. Its inverse image $T^{-n_0}\gamma_0$ is a subsegment of the HLUM $\gamma^u_{K(U_1)}(x)$. Since for typical x the parallelogram U_1 is dense, the image $T^{n_0}U_1$ has high density on γ_0. The map transforming $T^{n_0}x$ to $T^k\tilde{x}$ is a contraction of the HLUM γ_0 with coefficient close to $\Lambda_1 = \Lambda^u_k(y_0)$ (this is the coefficient of contraction of the LUM $\gamma^u(y_0)$ at y_0 under T^{-k}). Since U_0 and U_1 are 0-homogeneous, the conditional measures of U_1 and ΦU_1 on $T^{-n_0}\gamma_0$ are in the ratio $1 : (\text{const } \Lambda_1^{-1})$. This leads to the required estimate: $\nu(\Phi U_1) \geqslant \text{const } \Lambda_1^{-1}\nu(U_1)$. Similarly, $\nu(\Phi_1 U_1) \geqslant \text{const } \Lambda_2^{-1}\nu(U_1)$, where $\Lambda_2 = \Lambda^s_k(y_0)$ is the coefficient of contraction of the LSM $\gamma^s(y_0)$ at y_0 under T^k.

Theorem 1.6 has been proved.

Appendix 1

The proofs of Theorem 3.6 and Lemma 3.7 rest on an intricate and rather difficult study of the singularities of the map T. We present this study here in several stages, by considering in succession the various classes of billiards for which Theorems 1.1 and 1.2 were formulated above.

A1.1. Preparatory estimates for scattering billiards.

The quantities appearing in Definition 3.1 and Lemma 3.7 are strongly related to one another, therefore it is convenient to estimate them simultaneously. Further, the duality principle (replace T by T^{-1} and, correspondingly, γ^s by γ^u, B^s by B^u, and so on, for more detail see [4]) allows us to reduce Theorem 3.6 and Lemma 3.7 to the single proposition:

***Proposition* A1.1.** *Let x, y be points on a single n-homogeneous* LSM. *Then*

a) $|B^u(x)/B^u(y) - 1| \leqslant C\alpha_0^n$;

b) $|B^s(x)/B^s(y) - 1| \leqslant C\alpha_0^n$;

c) $|J^u(x, y) - 1| \leqslant C\alpha_0^n$;

d) $|\Lambda^s_k(x)/\Lambda^s_k(y) - 1| \leqslant C\alpha_0^n$ *for any* $k \geqslant 1$.

Here $J^u(x, y)$ is the Jacobian of the canonical isomorphism of the LUM $\gamma^u(x)$ and $\gamma^u(y)$ with respect to the measure $p(\cdot)$ (at x), and $\Lambda_k^s(x)$ is the local coefficient of contraction of the LSM $\gamma^s(x)$ at x under the action of T^k.

Remark. In Proposition A1.1, and in our propositions below, the symbol C denotes various constant quantities, determined by the choice of n_0 and ν in Definition 3.3 (and, of course, depending on the billiard domain Q), whose precise values are not of interest to us.

We recall (§2) that the lengths of all LUM and LSM, and the coefficients of expansion and contraction, are determined, unless otherwise stated, with respect to the measure p.

Proof. We begin by reducing assertions c) and d) to assertions a) and b). The general properties of T (§2) imply that $\Lambda_k^s(x) = \lambda^s(x)\cdot\ldots\cdot\lambda^s(T^{k-1}x)$, where $\lambda^s(x) = 1+\tau(x)B_-^s(Tx)$ and $B_-^s(x) = R(Tx)+B^s(Tx)$. Hence,

$$\text{(A1.1)}\qquad \frac{\Lambda_k^s(x)}{\Lambda_k^s(y)} = \prod_{i=1}^{k} \frac{1+\tau(T^{i-1}x)\,B_-^s(T^ix)}{1+\tau(T^{i-1}y)\,B_-^s(T^iy)}.$$

Similarly, the coefficient of expansion of the LUM $\gamma^u(x)$ at x under the action of T_k, $k \geqslant 0$, can be expressed by the formula $\Lambda_k^u(x) = \lambda^u(x)\cdot\ldots\cdot\lambda^u(T^{k-1}x)$, where $\lambda^u(x) = 1+\tau(x)B^u(x)$. As was proved in [2], [22], [27],

$$\text{(A1.2)}\qquad J^u(x, y) = \lim_{k\to\infty}\prod_{i=0}^{k} \frac{1+\tau(T^ix)\,B^u(T^ix)}{1+\tau(T^iy)\,B^u(T^iy)}.$$

From (A1.1) and (A1.2) it is clear that for proving c) and d) it suffices to estimate the nearness of the quantities $B^{u,s}(T^ix)$ and $B^{u,s}(T^iy)$, as well as of the quantities $\tau(T^ix)$ and $\tau(T^iy)$. More precisely, c) and d) follow from a) and b) and the single additional lemma:

***Lemma* A1.2.** *Let x, y be points on a single n-homogeneous* LSM *or* LUM. *Then*

$$\frac{|\tau(x)-\tau(y)|}{1+\tau(y)\,B^u(y)}\,B^u(x) \leqslant C\alpha_0^n.$$

For the time being we leave aside the proof of this lemma (it is given in §A1.2, §A1.3), and turn to assertions a) and b). To compare two continued fractions ($B^{u,s}(x)$ and $B^{u,s}(y)$, see §2) there is a method which we will demonstrate by the examples of the continued fractions $B_-^s(x)$ and $B_-^s(y)$.

We recall (§2) that $B_-^s(x) = R(x)+B^s(x) = R(x)+(\tau(x)+B_-^s(Tx))^{-1}$. Hence

$$B_-^s(x)-B_-^s(y) = R(x)-R(y)+B^s(x)\,B^s(y)\,((B^s(y))^{-1}-(B^s(x))^{-1}) =$$

$$= R(x)-R(y)-\frac{B_-^s(Tx)\,B_-^s(Ty)}{\lambda^s(x)\,\lambda^s(y)}\,(\tau(x)-\tau(y))+\frac{1}{\lambda^s(x)\,\lambda^s(y)}\,(B_-^s(Tx)-B_-^s(Ty)).$$

Repeating this expansion k times in succession, we obtain

$$\text{(A1.3)} \qquad B^s_-(x) - B^s_-(y) = \sum_{i=0}^{k-1} (\Delta R)_i - \sum_{i=0}^{k-1} (\Delta\tau)_i + (\Delta B)_k,$$

where we have put

$$(\Delta R)_i = \frac{1}{\Lambda^s_i(x)\,\Lambda^s_i(y)} (R(T^i x) - R(T^i y)),$$

$$(\Delta\tau)_i = \frac{B^s_-(T^{i+1}x)\,B^s_-(T^{i+1}y)}{\Lambda^s_{i+1}(x)\,\Lambda^s_{i+1}(y)} (\tau(T^i x) - \tau(T^i y)),$$

$$(\Delta B)_k = \frac{1}{\Lambda^s_k(x)\,\Lambda^s_k(y)} (B^s_-(T^k x) - B^s_-(T^k y)).$$

A similar expansion holds for $B^u(x) - B^u(y)$, but it is constructed from k reflections "in the past" (that is, from the points $T^{-1}x$, $T^{-2}x$, ..., $T^{-k}x$). We note that if $k \leqslant n$, then the points $T^{-k}x$ and $T^{-k}y$ lie on a single HLSM.

Remark 1. Since T is hyperbolic (see §2.2), $|\Lambda^{u,s}_i(x)| \geqslant \Lambda_0^{[i/m_0]}$ for all $x \in M$ and $i \geqslant 1$. Moreover, the lengths of n-homogeneous LUM and LSM do not exceed $C\lambda_0^n$, where $\lambda_0 = \Lambda_0^{-1/m_0}$.

Remark 2. For any $x \in M$ and $i \geqslant 1$ we have $B^s(x)\Lambda^s_i(x) \geqslant B^s_-(T^i x)$, which can easily be verified using the representation of $B^s(x)$ as a continued fraction (2.5).

The expansion (A1.3) and these remarks reduce assertions a) and b) to Lemma A1.2 and the following two:

***Lemma* A1.3.** *If the points x, y lie on a single n-homogeneous* LSM *or* LUM, *then*

$$\frac{|R(x) - R(y)|}{B^u(x)} \leqslant C\alpha_0^n.$$

***Lemma* A1.4.** *If the points x, y lie on a single* HLUM *or* HLSM, *then*

$$\frac{B^u(x)}{B^u(y)} \leqslant C.$$

In Lemmas A1.2–A1.4 there figure two points x, y on a single LUM or LSM. In the sequel we denote by γ_0 the part of this LUM or LSM bounded by the points x, y.

It remains to prove Lemmas A1.2–A1.4. We first give their proofs for relatively simple billiards, and then for more complicated ones. Here, by simple we mean billiards whose boundary ∂Q does not have break points (this is possible when Q is a torus with several cut-out convex "holes" having a smooth boundary). In this case singularities of T arise only for tangential reflections, that is, on the set $S_0 \cup S_{-1}$ (since V_0 is empty). The break points of ∂Q will add another kind of singularity: on the set $V_0 \cup V_{-1}$.

A1.2. Scattering billiards without break points on the boundary.
In this case the time of free movement is bounded away from zero: $\tau(x) \geqslant \tau_0 > 0$. Hence the inequality in Lemma A1.2 reduces to the simpler

$$\text{(A1.4)} \qquad |\tau(x) - \tau(y)| \leqslant C\alpha_0^n.$$

For proving the latter we use the obvious relation

$$\text{(A1.5)} \quad |\tau(x) - \tau(y)| \leqslant |r(x) - r(y)| + |r(Tx) - r(Ty)|.$$

The quantity on the right-hand side of (A1.5) does not exceed $C(\sqrt{p(\gamma_0)} + \sqrt{p(T\gamma_0)})$, see (2.3) (here we must pay attention to the fact that if γ_0 is an LUM and $n = 0$, then its image $T\gamma_0$ lies entirely on one regular component of ∂Q, in correspondence with the definition of HLUM, see §3.3). Now Lemma A1.2 follows from the hyperbolicity of T (see (2.2)).

The restriction $\tau_-(x) \geqslant \tau_0 > 0$ implies that

$$\text{(A1.6)} \qquad |B^u(x) - R(x)| \leqslant C.$$

The function $R(x) = 2\varkappa(x)/\cos\varphi(x)$ is continuously differentiable on M, and is infinite on S_0. Our subdividing segments (§3) are constructed so that $R(x)$ does not undergo too large oscillations inside the strips between these segments (the latter will be called *homogeneity strips* for short). Since γ_0 is homogeneous, it lies entirely in one homogeneity strip, for example, with index k (that is, in the strip given by the inequalities $(k+1)^{-\nu} < \pi/2 - |\varphi| < k^{-\nu}$). It is then easily seen that $\cos\varphi(x) \approx \cos\varphi(y) \approx Ck^{-\nu}$, hence $R(x) \approx R(y) \approx Ck^{\nu}$. This and (A1.6) imply Lemma 1.4 and also, by (2.4),

$$\text{(A1.7)} \qquad 0 < C_1 \leqslant \left| \frac{\varphi(x) - \varphi(y)}{r(x) - r(y)} \right| \leqslant C_2 < \infty,$$

where C_1 and C_2 depend only on the domain Q. The last inequality implies that $|\varphi(x) - \varphi(y)| \leqslant Ck^{1-\nu}$, and hence

$$p(\gamma_0) \approx C\,|r(x) - r(y)|\cos\varphi(x) \approx C\,|\varphi(x) - \varphi(y)|\,k^{-\nu}.$$

From this we can easily derive two estimates which we will need in the sequel:

$$\text{(A1.8)} \qquad |r(x) - r(y)| \leqslant (p(\gamma_0))^{\frac{1+\nu}{1+2\nu}},$$

$$\text{(A1.9)} \qquad (p(\gamma_0))^{\frac{\nu}{1+2\nu}} \leqslant (B^u(x))^{-1} \approx C/R(x).$$

Now we prove Lemma A1.3 as follows:

$$\left| \frac{R(x) - R(y)}{B^u(x)} \right| \leqslant C \left| \frac{\varphi(x) - \varphi(y)}{\cos\varphi(x)} \right| \leqslant C\,(p(\gamma_0))^{\frac{1}{1+2\nu}}$$

and, finally, we use the hyperbolicity of T.

We note that inequalities (A1.8) and (A1.9) are true also outside a neighbourhood of S_0, that is, for *any* HLUM (HLSM) γ_0 with end-points at x and y. In the following subsection we prove that they hold for scattering billiards with break points on the boundary.

A1.3. Scattering billiards with break points on the boundary.

We immediately note that in this case the derivation of Lemmas A1.2–A1.4 is substantially more complicated. In essence, up to now nobody has undertaken the analysis of the singularities of T on the set $V_0 \cup V_{-1}$, since the authors of studies on the ergodic properties of billiards have usually excluded those cases from consideration ([13], [5], [22], [27]).

Fortunately, in these cases there is one circumstance which essentially simplifies the matter:

***Lemma* A1.5.** *For any scattering billiard satisfying condition* A *there are an* $\varepsilon_0 > 0$ *and a* $\varphi_0 > 0$ *such that in any series of successive reflections the trajectories of the billiard starting in an* ε_0*-neighbourhood of a break point of the boundary* ∂Q *there can be only one "almost tangential" reflection, in which the angle of reflection* $> \pi/2 - \varphi_0$. *If there is such a reflection, then it is extremal (first or last) in the series under consideration, that is, before or after it the length of a free path* $> \varepsilon_0$.

This lemma means that "double trouble", when $\cos\varphi(x)$ and $\tau_\pm(x)$ are simultaneously close to zero, does not happen often: *at most once* in every series of reflections starting in a small neighbourhood of a "billiard corner".

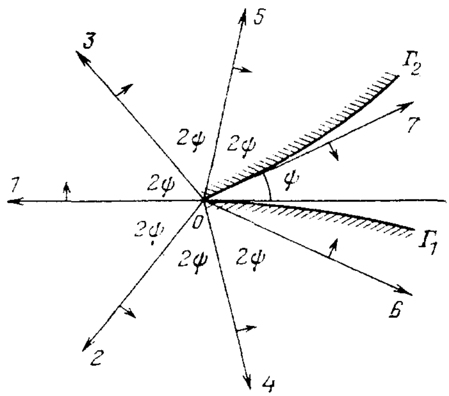

Fig. 7

Proof of Lemma A1.5. The boundary ∂Q has finitely many break points, hence it suffices to consider one of these only. In this form the lemma becomes a geometrical problem of almost High School level. The key to its solution is contained in Fig. 7. In this figure the components Γ_1 and Γ_2 of ∂Q form an angle ψ at the vertex O. If ε_0 is sufficiently small and the trajectory is "almost" tangent to Γ_1 at the first reflection, then the next reflection "rotates" its velocity vector as indicated in the figure; the index of the vector corresponds to the number of the reflection. According to this scheme, the second "almost" tangential reflection can occur either from Γ_1, if $\psi \approx \pi/n$ (n an integer), or from Γ_2, if $\psi \approx 2\pi/(2n+1)$ (n an integer). In the figure we have drawn the second possibility. Because of the bending of the boundaries Γ_1 and Γ_2, the actual positions of the velocity vectors differ somewhat from those drawn in the figure: they must be rotated around O in directions indicated by the short arrows over certain, although small, *positive* angles. These rotations also exclude the possibility of a repeated "almost" tangential reflection, since the last velocity vector will point to the strict interior of Q, which proves the lemma.

Remark A1.6. We recall that in the definition of the subdividing curves $\mathscr{D}$ in §3, to the subdividing segments $\mathscr{D}_0$ certain of their images were adjoined. Here we determine these images precisely: $\mathscr{D}_1 = \{T^i z : z \in \mathscr{D}_0$ and if $i > 0$, then $\tau(z) \leqslant \varepsilon_0$, $\tau(Tz) \leqslant \varepsilon_0$, ..., $\tau(T^{i-1}z) \leqslant \varepsilon_0$, while if $i < 0$, then $\tau(T^i z) \leqslant \varepsilon_0$, ..., $\tau(T^{-1}z) \leqslant \varepsilon_0\}$.

Remark. The quantities ε_0 and φ_0 depend only on the domain Q, hence in the sequel all constants denoted by C depend also on ε_0 and φ_0.

We turn to the proof of Lemmas A1.2–A1.4 in the case under consideration. If $\tau_\pm(x) > \varepsilon_0$ and $\tau_\pm(y) > \varepsilon_0$, then the proof can be given as in §A1.2. Suppose that at least one of the quantities $\tau_\pm(x)$ or $\tau_\pm(y)$ does not exceed ε_0, that is, we are in a small neighbourhood of a "billiard corner". Then there is an $i_0 \geqslant 0$ such that $\tau_-(T^{-i_0}x) > \varepsilon_0$ and $\tau_-(T^{-i_0}y) > \varepsilon_0$, but for all $0 \leqslant i < i_0$ either $\tau_-(T^{-i}x) \leqslant \varepsilon_0$ or $\tau_-(T^{-i}y) \leqslant \varepsilon_0$. We call the series of reflections with indices from $-i_0$ to 0 a *series of corner reflections*. If there is no "almost" tangential reflection in this series, that is, $|\varphi(T^i x)| \leqslant \pi/2-\varphi_0$ for all $-i_0 \leqslant i \leqslant 0$, then $B^u(x) \leqslant C < \infty$. This follows easily from the estimate

$$B^u(x) \leqslant R(x) + R(T^{-1}x) + \ldots + R(T^{-i_0}x) + (\tau_-(T^{-i_0}x))^{-1} \tag{A1.10}$$

and the restriction $i_0 \leqslant m_0$ (see Remark 2.1). The rest of the proof of these two assertions is as in §A1.2.

Suppose now that there is one "almost" tangential reflection in the series of corner reflections under consideration, that is $|\varphi(T^j x)| \geqslant \pi/2-\varphi_0$ for some $j \in [-i_0, 0]$. By Lemma A1.5 this reflection is extremal, that is, either $j = -i_0$ or $j = 0$. We note that in both cases the point $T^j x$ belongs to a neighbourhood of $S_0 \cap V_0$, that is, it is near a vertex of one of the rectangles making up the phase space $\overline{M}$.

Using (A1.10) it is easily proved that an analogue of (A1.6) holds at T^jx: $|B^u(T^jx)-R(Tx)| \leqslant C$. This and (2.4) imply that

$$0<C_1\leqslant\left|\frac{\varphi(T^jx)-\varphi(T^jy)}{r(T^jx)-r(T^jy)}\right|\leqslant C_2<\infty.$$

We note that in this way the property (A1.7) holds for all HLUM and HLSM in some neighbourhood of S_0 (in §2 we noted that in a neighbourhood of V_0 it is sometimes violated). Below we consider the cases $j = 0$ and $j = -i_0$ separately.

The case $j = 0$ is relatively simple. By Lemma A1.5, $\tau(x) > \varepsilon_0$ and $\tau(y) > \varepsilon_0$, and by (A1.10) relation (A1.6) holds, and hence (A1.8) and (A1.9) hold. Hence Lemmas A1.2–A1.4 can be proved as in §A1.2.

The case $j = -i_0$ requires a more refined analysis. The fact is that in this case the quantity $B^u(x)$ can also become arbitrarily large, but not because of a large $R(x)$, but because of a large $R(T^{-i_0}x)$ and small "transition times" $\tau(T^{-i_0}x), \ldots, \tau(T^{-1}x)$. Pictorially speaking, the LUM $\gamma^u(T^{-i_0}x)$ does not manage to "become straightened" by an "almost" tangential reflection at $T^{-i_0}x$, hence its curvature at x remains large. In this case the proof of Lemmas A1.2–A1.4 is based on a number of technical estimates.

First, the quantities $R(x), \ldots, R(T^{-i_0+1}x)$ in (A1.10) are uniformly bounded above, and since $i_0 \leqslant m_0$ we obtain a bound on $B^u(x)$:

$$B^u(x)\leqslant B^u(T^{-i_0}x)+C \tag{A1.11}$$

We now prove the inequality

$$|\tau(T^ix)-\tau(T^iy)|\leqslant C(p(T^{-i_0}\gamma_0))^{\frac{1+\nu}{1+2\nu}} \tag{A1.12}$$

for all $i = -i_0, -i_0+1, \ldots, -1$, and also for $i = 0$ provided that $\tau(y) \leqslant \varepsilon_0$.

We consider the various cases. Let $i = -i_0$. Then the curve $T^i\gamma_0$ lies in a neighbourhood of S_0, that is, in one of the homogeneity strips, and the estimates (A1.6)–(A1.9) hold for the points x and y. In this case the estimate (A1.12) can be easily derived using (A1.5).

Let $i > -i_0$. Then the curves $T^i\gamma_0$ and $T^{i+1}\gamma_0$ lie outside a neighbourhood of S_0, and, by the same token, $p(T^i\gamma_0) \geqslant C|r(T^ix)-r(T^iy)|$ and $p(T^{i+1}\gamma_0) \geqslant C|r(T^{i+1}x)-r(T^{i+1}y)|$. By (A1.5) we obtain

$$|\tau(T^ix)-\tau(T^iy)|\leqslant C(p(T^i\gamma_0)+p(T^{i+1}\gamma_0)). \tag{A1.13}$$

If γ_0 is an LSM, then (A1.12) is immediately obtained from (A1.13). If γ_0 is an LUM, then in addition we have to estimate the coefficient of expansion $\Lambda^u_l(z)$ at each point z of the LUM $T^{-i_0}\gamma_0$ under the action of T^l, where $l = 1, 2, \ldots, i_0+1$. We show that

$$\Lambda^u_l(z)\leqslant C(1+(\tau(z)+\ldots+\tau(T^{l-1}z))B^u(z)). \tag{A1.14}$$

This means that (up to a factor C) $\Delta^u_\gamma(z)$ is determined only by the principal ("almost" tangential) reflection at z and the time of movement on the LUM from z to $T^l z$.

For $l = 1$ formula (A1.14) is obvious. We prove it for $l = 2$:

$$(1 + \tau(z) B^u(z))(1 + \tau(Tz) B^u(Tz)) =$$
$$= (1 + (\tau(z) + \tau(Tz)) B^u(z) + R(Tz)\tau(Tz)\ (1 + \tau(z) B^u(z)) \leqslant$$
$$\leqslant (1 + R(Tz)\tau(Tz))\ (1 + (\tau(z) + \tau(Tz)) B^u(z)).$$

For $l \geqslant 3$ the estimate (A1.14) can be proved by induction, taking into account the restriction $l \leqslant m_0$.

Remark. In the proof of (A1.14) and in the sequel, the fact that if γ_0 is an LUM, then its inverse image $T^{-i_0}\gamma_0$ must belong entirely to one of the homogeneity strips, is essential for us. Precisely for this we introduced in §3 the additional subdividing curves $\mathscr{D}_1$.

The estimate (A1.14) just proved implies that in (A1.13) we have $p(T^i\gamma_0) \leqslant p(T^{i+1}\gamma_0) \leqslant CB^u_{\max}(T^{-i_0}\gamma_0)p(T^{-i_0}\gamma_0)$, where $B^u_{\max}(T^{-i_0}\gamma_0)$ denotes the maximum of B^u on the curve $T^{-i_0}\gamma_0$. Finally, this and (A1.9) imply (A1.12).

Now we prove Lemmas A1.2–A1.4. We begin with the last one. For $i_0 = 0$ it follows from (A1.6). For $i_0 = 1$ we have

$$\frac{B^u(x)}{B^u(y)} = \frac{R(x) + (\tau(T^{-1}x) + (B^u(T^{-1}x))^{-1})^{-1}}{R(y) + (\tau(T^{-1}y) + (B^u(T^{-1}y))^{-1})^{-1}} \leqslant$$
$$\leqslant C(1 + |\tau(T^{-1}x) - \tau(T^{-1}y)|\, B^u(T^{-1}y))$$

and we subsequently use the already proved formulae (A1.11) and (A1.12). For $i_0 \geqslant 2$ the lemma can be proved by induction.

Lemma A1.3 for $i_0 = 0$ can be proved as in §A1.2, and for $i_0 \geqslant 1$ the curve γ_0 lies outside a neighbourhood of S_0, whence $|R(x) - R(y)| \leqslant C(|r(x) - r(y)| + |\varphi(x) - \varphi(y)|) \leqslant C\sqrt{p(\gamma_0)}$ (see (2.3)).

The proof of Lemma A1.2 for $\tau(y) > \varepsilon_0$ goes through as in §A1.2. For $\tau(y) \leqslant \varepsilon_0$ it suffices to apply in succession (A1.12) (for $i = 0$), (A1.11), and (A1.9):

$$|\tau(x) - \tau(y)|\, B^u(x) \leqslant C(p(T^{-i_0}\gamma_0))^{\frac{1+\nu}{1+2\nu}} B^u(T^{-i_0}x) \leqslant C(p(T^{-i_0}\gamma_0))^{\frac{1}{1+2\nu}}.$$

Thus, Assertions 3.6 and 3.7 have been proved for all scattering billiards.

A1.4. Semiscattering billiards.

Our next tactic is to reduce the case of semiscattering and focussing billiards to that of scattering billiards, which has already been dealt with. This operation is easiest for semiscattering billiards. In this case the continued fractions $B^{u,s}(x)$ have vanishing even terms ($R(T^i x) = 0$) at the points corresponding to reflections in the neutral boundary components. The obvious relation

$$\tau_i + \cfrac{1}{0 + \cfrac{1}{\tau_{i+1} + \cdots}} = \tau_i + \tau_{i+1} + \cdots$$

allows us to "contract" the fractions $B^{u,s}(x)$, by excluding all zero terms. After this, the odd terms of the "contracted" fractions will correspond to intervals of movement *between successive reflections in the scattering boundary components* (the reflections in the neutral components are simply discarded).

Geometrically this means that instead of reflection of an LUM (LSM) in a neutral component Γ_0, we reflect the whole domain Q with respect to the component Γ_0, while the LUM (LSM) continues to move inside the reflected domain as if the wall Γ_0 were transparent (Fig. 8). Here, the geometrical characteristics of the initial LUM (LSM), that is, its curvature and length at each moment of time, will be the same as for this "double", drawn through Γ_0 without reflection. Therefore the properties of LUM and LSM in semiscattering billiards are the same as for their doubles, moving according to the laws of a scattering billiard. All computations made in §§A1.1 – A1.3 are exactly applicable to these doubles, hence all results can be carried through unchanged.

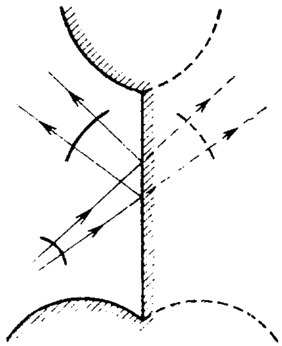

Fig. 8

The idea of doubles of LUM and LSM will also be used in the following subsection.

A1.5. Hyperbolic billiards with focussing boundary components.
First of all we note that in this case the continued fractions $B^{u,s}(x)$ are alternating, and hence many computations from §§A1.1 – A1.3 need to be made more precise. Moreover, the derived map T is defined differently from the case of scattering billiards (see §2), and we can "neutralize" this distinction by using the idea of doubles of LUM and LSM from the previous subsection.

First we modify the continued fractions $B^{u,s}(x)$. Suppose that the trajectory of the billiard experiences n successive reflections in a single focussing component $\Gamma_1 \subset \partial Q$. We call such a series of reflections a *Φ-series*. Since Γ_1 is an arc of a circle, all angles of reflection are equal ($=\varphi$), and the intervals of movement between reflections are constant ($\tau = 2R_1 \cos\varphi$, where R_1, here and below, denotes the radius of the arc Γ_1). The corresponding parts of the continued fraction $B^{u,s}(x)$ allow the following "contraction":

$$\text{(A1.15)}\qquad \left.\begin{array}{l} -\dfrac{4}{\tau} + \cfrac{1}{\tau + \cfrac{1}{-\dfrac{4}{\tau} + \cfrac{1}{\tau + \ddots + \cfrac{1}{-\dfrac{4}{\tau} + \cdots}}}} \\ n \text{ times} \end{array}\right. = -\frac{2}{\tau} + \cfrac{1}{-(n-1)\tau + \cfrac{1}{-\dfrac{2}{\tau} + \cdots}}.$$

Here, if $n = 1$, then on the right-hand side we have the same continued fraction $-\frac{4}{\tau} + \ldots$ as on the left. Inequality (A1.15) can be easily verified by induction.

The transformed fraction $B^u(x)$, in which all Φ-series are "contracted" by using (A1.15), corresponds to the motion of an LUM-double by the following laws. Upon approaching the successively focussing component Γ_1 of radius R_1, it reflects in a (focussing) arc of doubled radius $2R_1$, then moves backwards (!) during time $(n-1)\tau$ (that is, each time the amount of time which the original LUM spends between the first and the next reflection in Γ_1), then again reflects in the (focussing) arc of radius $2R_1$, and further moves as the original LUM. We see that the complete Φ-series is replaced by two reflections. As distinct from the previous subsection, our doubles have geometrical characteristics (curvature and length) that are identical with those of the original LUM and LSM (the "originals"), not at all moments of time, but only at the moments of reflection, from which the derived map T is constructed; clearly, this is sufficient for our purposes.

In connection with the construction of doubles of LUM and LSM, it is convenient to introduce a new derived map, $\hat{T}$, constructed from all reflections of these doubles. More precisely, $\hat{T}$ is the derived map of the billiard flow acting in the set $\hat{M} = M \cup \hat{M}_-$, where $\hat{M}_-$ combines all points of *first* reflection in Φ-series (in the coordinates r, φ, the set $\hat{M}_-$ can be regarded as a set of parallelograms which are symmetric to those depicted in Fig. 3). Although the counter of the number of reflections changes in this construction, it does so by at most a factor 2: to the point $T^n x$ will correspond $\hat{T}^{N(n)}x$ for some $N(n) \in [n, 2n]$; hence we have completely proved Theorem 3.6 and Lemma 3.7 for $\hat{T}$.

For each $x \in \hat{M}_-$ we introduce the notations $\hat{\tau}(x)$, $\hat{B}^{u,s}(x)$, $\hat{R}(x)$, $\hat{\gamma}^{u,s}(x)$, and $\hat{\lambda}^{u,s}(x)$, analogous to the corresponding notations for $x \in M$. The properties of these quantities are different on the three parts of the phase space $\hat{M}$:

1. If $x \in M_+$, then all properties are the same as in the case of scattering billiards.

2. If $x \in \hat{M}_- \setminus M_-$ (a *first* reflection in a Φ-series), then $\hat{R}(x) = -(R_1 \cos \varphi(x))^{-1} < 0$, $\hat{\tau}(x) = -2(n-1)R_1 \cos \varphi(x) < 0$ (where n is the number of reflections in the Φ-series), and $\lambda^u(x) = 1 + \hat{\tau}(x)\hat{B}^u(x) > 1$.

3. If $x \in M_-$ (a *last* reflection in Φ-series), then $\hat{R}(x) = -(R_1 \cos \varphi(x))^{-1} < 0$, $\hat{\tau}(x) = \tau(x) > 0$, and $\hat{B}^u(x)$ is given by formula (A1.15), in which $\tau = 2R_1 \cos \varphi$, that is, $\hat{B}^u(x) < 0$. By the defocussing condition (§2), $\hat{\tau}(x) = \hat{\tau}_\Phi(x) + \hat{\tau}_P(x)$, where $\hat{\tau}_\Phi(x)$ and $\hat{\tau}_P(x)$ are the intervals of movement up to and after the focussing point, and $\hat{\tau}_\Phi(x) < \hat{\tau}_P(x)$. We call the difference $\hat{\tau}_E(x) = \hat{\tau}_P(x) - \hat{\tau}_\Phi(x)$ the *effective time interval* between reflections at the points $x \in M$ and $\hat{T}x$. It is easily verified that

$$\text{(A1.16)} \quad \hat{\lambda}^u(x) = |1 + \hat{\tau}(x)\hat{B}^u(x)| = 1 + \hat{\tau}_E(x)\,|\hat{B}^u(x)| > 1.$$

Similar formulae hold for LSM. We see that the expressions for $\hat{\lambda}^{u,s}(x)$ differ from those of scattering billiards only by the appearance of an absolute value in one of the three cases. Therefore all computations from §A1.1 remain valid also in the present situation, and in the expressions for $\lambda^{u,s}$ and $B^{u,s}$ the symbol for absolute value appears. It remains to discuss in more detail the two remarks given before Lemma A1.3, but first we prove two useful assertions.

***Lemma* A1.7.** *The lengths of effective time intervals are bounded away from zero*: $\hat{\tau}_E(x) \geqslant \hat{\tau}_0 > 0$ *for all* $x \in M_-$.

In essence, this is a purely geometrical assertion. To prove it we assume that $\hat{\tau}_E(x_i) \to 0$ for certain $x_i \in M_-$. If $\sup |\varphi(x_i)| = \pi/2$, then some subsequence of points x_{i_k} converges to the tangent drawn at an end-point of one of the focussing components. By condition F2 (§2), $\tau(x_{i_k}) \to \text{const} > 0$, while $\tau_\Phi(x_{i_k}) \to 0$; we have obtained a contradiction. If, on the other hand, $\sup |\varphi(x_i)| < \pi/2$, then, using the compactness of M_-, we can extract a convergent subsequence $x_{i_k} \to x_0$. Then $|\varphi(x_0)| < \pi/2$, and, moreover, $\tau_\Phi(x_0) = \tau_P(x_0)$. This is only possible if x_0 is a periodic point whose trajectory is reflected in only one focussing component; this contradicts condition F3. The lemma has been proved.

Remark A1.8. $\hat{\tau}(x) \geqslant \tau_1(Q) > 0$ for all points $x \in M_+$ such that $\hat{T}x \in M_-$; moreover, $|\hat{\tau}(x)| \geqslant \tau_2(Q) > 0$ for all $x \in \hat{M}_- \setminus M_-$.

Assertions A1.7 and A1.8 imply the first remark (on hyperbolicity) given before Lemma A1.3 in §A1.1. To prove the second we again have to represent $B^s(x)$ as a continued fraction (2.5) and consider the three cases outlined above. In the second and third cases the required estimate follows from the additional relation $2B^s(x) \leqslant |\hat{R}(x)| = 2(R_1 \cos \varphi(x))^{-1}$ for $x \in M_-$, which, in turn, follows from the defocussing condition (§2).

We turn to the proof of Lemmas A1.2–A1.4 for the billiards considered here. For $x, y \in M_+$ and $\hat{T}x, \hat{T}y \in M_+$, it can be taken over from §§A1.2–A1.3 unchanged. If $x, y \in M_+$ but $\hat{T}x, \hat{T}y \in M_-$, only in the proof of Lemma A1.2 does there arise a new ingredient: it is necessary to estimate the difference $|r(\hat{T}x)-r(\hat{T}y)|$. For this we use inequality (2.6), from which we find that $|r(\hat{T}x)-r(\hat{T}y)| \leqslant \sqrt{p(\hat{T}\gamma_0)}$, as in §§A1.2–A1.3.

Now let $x, y \in M_- \cup \hat{M}_-$. Then by (A1.6) and Lemma A1.7 the inequality in Lemma A1.2 can be written in the simpler form

$$|\hat{\tau}(x) - \hat{\tau}(y)| \leqslant C\alpha_0^n, \tag{A1.17}$$

as already done in §A1.2.

If $x, y \in M_-$ (that is, x is the *last* reflection in a Φ-series), then (A1.17) can be proved using (A1.5) and (2.6). If, on the other hand, $x \in \hat{M}_- \setminus M_-$ (that is, x is the *first* reflection in a Φ-series), then $\hat{\tau}(x) = -2n(x)R_1 \cos\varphi(x)$ and $\hat{\tau}(y) = -2n(y)R_1 \cos\varphi(y)$, where $n(x)$ denotes the number of reflections following the point x in the Φ-series. Clearly, $n(x) = n(y) \leqslant C/\cos\varphi(x)$, and therefore (A1.17) follows from (2.7).

To prove Lemmas A1.3 and 1.4 we have to estimate $\hat{R}(x)$. As in the case of scattering billiards, it is infinite on the set S_0 (for $\cos\varphi(x) = 0$). However, in this case it is not necessary to partition a neighbourhood of S_0 into homogeneity strips, since it has already been partitioned in a necessary manner by the discontinuity curves (see §2) (these curves correspond approximately to subdividing segments with exponent $\nu = 1$).

On the other hand, for $x \in M_- \cup \hat{M}_-$ we only have to prove the single additional estimate

$$|\hat{B}^u(x)| \geqslant C/\cos\varphi(x) \tag{A1.18}$$

for some $C > 0$, which is obvious for scattering billiards.

If $x \in M_-$ (the *last* reflection in a Φ-series), then by the defocussing condition $|\hat{B}^u(x)| \geqslant (R_1\cos\varphi(x))^{-1}$. If, however, $x \in \hat{M}_-\setminus M_-$ (the *first* reflection in a Φ-series), then by (A1.15)

$$\hat{B}^u(x) = -\frac{1}{R_1\cos\varphi(x)} + \frac{1}{\tau(\hat{T}^{-1}x) + (\hat{B}^u(\hat{T}^{-1}x))^{-1}}.$$

By the defocussing condition, $\hat{\tau}(\hat{T}^{-1}x) \geqslant 2R_1\cos\varphi(x)$ and $\hat{B}^u(\hat{T}^{-1}x) \geqslant \tau(\hat{T}^{-1}x)$, hence $|\hat{B}^u(x)|$ can only be close to zero if $\hat{T}^{-1}(x) \in M_-$, and then $\hat{\tau}(\hat{T}^{-1}x) \approx 2R_1\cos\varphi(x)$ and $\hat{\tau}_P(\hat{T}^{-1}x) \approx \hat{\tau}_\Phi(\hat{T}^{-1}x)$. In view of the condition F2 (§2) this cannot hold, which proves (A1.18).

The estimates (A1.18) and (2.7) imply that

$$\left|\frac{\hat{R}(x)-\hat{R}(y)}{\hat{B}^u(x)}\right| \leqslant C\left|\frac{\varphi(x)-\varphi(y)}{\cos\varphi(y)}\right| \leqslant C(p(\gamma_0))^{1/3},$$

which proves Lemma A1.3. It also implies Lemma A1.4, which ends the analysis of billiards with focussing components.

Appendix 2

Basically, our proof of Theorem 3.8 follows the scheme for constructing LUM and LSM in [13].

For an arbitrary $\varepsilon > 0$ and subset $A \subset M$ we denote by $U_\varepsilon(A)$ the neighbourhood of A formed by all 1-increasing curves of length $\leqslant \varepsilon$ intersecting A.

For scattering billiards with finite horizon the number of discontinuity curves in $R_{-1,1}$ is finite, and hence $\nu(U_\varepsilon(R_{-1,1})) \leqslant$ const ε. When there appear singular points of infinite horizon and sliding types, an analysis of the structure of the discontinuity curves in neighbourhoods of these points ([8], see §2) allows us to obtain the estimate $\nu(U_\varepsilon(R_{-1,1})) \leqslant \text{const } \varepsilon^{4/5}$.

It is also not difficult to show that $\nu(U_\varepsilon(\mathscr{D})) \leqslant \text{const } \varepsilon^{2\nu/(1+4\nu)}$. In fact, the construction of the subdividing segments $\mathscr{D}_0$ easily implies the required estimate for $\nu(U_\varepsilon(\mathscr{D}_0))$. To extend it to the additional subdividing curves $\mathscr{D}_1$, we consider an arbitrary 1-increasing curve γ of length $\leqslant \varepsilon$ intersecting $\mathscr{D}_1$. Then for some $|i| \leqslant m_0$ the curve $T^i\gamma$ intersects $\mathscr{D}_0$. By Lemma A1.5, for $x \in \gamma$ we have $B^u(x) \leqslant$const(Q), and hence $p(T^u\gamma) \leqslant$ const ε. This then implies the required estimate.

Further, we choose $p_0 > 0$, put $\varepsilon_n = p_0\Lambda_0^{-[n/m_0]}$ for all $n \geqslant 0$, and note that

$$\text{(A2.1)} \qquad \sum_{n=0}^{\infty} \nu\left(U_{\varepsilon_n}\left(R_{-1,1} \cup \mathscr{D}\right)\right) \leqslant \text{const } (Q)\, p_0^d,$$

where $d = \min\{4/5,\ 2\nu/(1+4\nu)\}$.

Standard reasonings [13], [29] imply that if $T^n x \notin U_{\varepsilon_n}(R_{-1,1} \cup \mathscr{D})$ for all integers $n \geqslant 0$, then there is an HLUM $\gamma^{0u}(x)$, and the distance from its end-points to x is at least p_0. By (A2.1) the measure of the set of such points, for fixed p_0, is at least $1-\text{const}(Q)p_0^d$. Thereby Theorems 3.8 and 3.10 have been established.

To prove Proposition 3.11 we consider a point x such that on the LUM $\gamma^u(x)$ the subdividing points (that is, the points in the intersections $T^n\mathscr{D} \cap \gamma^u(x)$ for all $n \geqslant 0$) have a limit point somewhere inside the curve $\gamma^u(x)$. Then for infintely many $n \geqslant 0$ the relation $T^{-n}x \in U_{\varepsilon_n}(\mathscr{D})$ holds, where $\varepsilon_n = p(\gamma^u(x))\Lambda_0^{-[n/m_0]}$. By (A2.1) and the Borel–Cantelli lemma such points x form a set of measure zero.

Appendix 3

First we recall the analogue of Theorem 3.12 proved in [24], given here in somewhat strengthened form:

***Theorem* A3.1.** *Assume that a scattering billiard satisfies the conditions* A, B, C (§2) *and has finite horizon* ($\tau(x) \leqslant$ const $< \infty$). *Let* γ^u *be an arbitrary* LUM *and* $D > 0$. *For each* $N \geqslant 0$ *we choose on* γ^u *subsegments* $\gamma^u_{N,i}$, *being the* *smoothness components of the map T^N on γ^u and such that for all $n \in [0, N]$ the image $T^n\gamma^u_{N,i}$ lies in a smoothness component of the set $T^N\gamma^u$ that has length at most D. Then there are numbers $D > 0$, $C > 0$, and $0 < \lambda < 1$ such that for all $N \geqslant 1$*

$$p(\bigcup_i \gamma^u_{N,i}) \leqslant C\lambda^N.$$

In other words, the total measure of the set of points on the LUM γ^u whose images "in the future" during the first N steps do not fall once in a sufficiently long LUM (that is, in an LUM of length $> D$) decays exponentially with N.

For the proof we first write down the obvious estimate

$$\text{(A3.1)} \qquad p(\bigcup_i \gamma^u_{N,i}) \leqslant D\sum_i \Lambda_N^{-1}(i),$$

where $\Lambda_N(i)$ is a (lower) estimate for the coefficient of expansion of the LUM $\gamma^u_{N,i}$ under the action of T^N.

Since T is hyperbolic (§2), we may put $\Lambda_N(i) = \Lambda_0^{[N/m_0]}$. It remains to estimate the number of segments $\gamma^u_{N,i}$ for a given N. Let $m \geqslant 1$ be an integer. Condition B (§2) implies that if D is sufficiently small ($D \leqslant D_0(m)$), then any LUM of length $\leqslant D$ intersects at most K_0m discontinuity curves in $R_{-m,0}$. Hence the number of segments $\gamma^u_{N,i}$ does not exceed $(K_0m)^{[N/m]+1}$.

Together with (A3.1) this estimate gives

$$p(\bigcup_i \gamma^u_{N,i}) \leqslant D\Lambda_0^{-[N/m_0]}(K_0m)^{[N/m]+1}.$$

We choose an m such that the quantity $\Lambda_1 = \Lambda_0(K_0m)^{-m_0/m}$ is larger than 1. Then

$$p(\bigcup_i \gamma^u_{N,i}) \leqslant \text{const}\, D\Lambda_1^{-[N/m_0]},$$

and Theorem A3.1 has been proved.

We turn to our Theorem 3.12. By analogy with Theorem A3.1, we denote by $\gamma^u_{N,i}$ all subsegments of γ^u whose images under T^n for all $0 \leqslant n \leqslant N$ fall within the homogeneous components of the set $T^n\gamma^u$ having length at most D. Then the estimate (A3.1) remains valid.

However, in our case the number of segments $\gamma^u_{N,i}$ can be infinite, and for various reasons: subdivision in a neighbourhood of S_0 and accumulation of infinitely many discontinuity curves in a neighbourhood of the singular points, both of infinite horizon type (§2.3) and of sliding type (§2.5). On the other hand, in the parts of the phase space described above the coefficient of expansion of the LUM grows to infinity. This allows us to obtain a more precise estimate for $\Lambda_N(i)$, and thus to compensate for the growth of the number of segments $\gamma^u_{N,i}$.

We consider in more detail the two possible sources for the unbounded growth of the number of components $\gamma^u_{N,i}$.

1st case (subdivision). We enumerate all homogeneity strips (see Appendix 1) so that the n-th strip is given by the relation $\pi/2-n^{-\nu} \leqslant |\varphi| \leqslant \pi/2-(n+1)^{-\nu}$ $(n \geqslant n_0)$. For the time being we will only consider the strips with indices $\geqslant n_1$, where n_1 is a sufficiently large number, to be chosen below. We will consider the remaining strips (with indices from n_0 to n_1) below, along with the usual discontinuity curves. Suppose that an arbitrary LUM γ^u lies in the n-th homogeneity strip. Then the relations given in §§A1.2–A1.3 easily imply that the coefficient of expansion of this LUM under T^i, for some $i \leqslant m_0$, is at least const n^ν, and Lemma A1.5 implies that the intermediate images $T\gamma^u$, ..., $T^{i-1}\gamma^u$ cannot fall even once in the homogeneity strip.

2nd case (discontinuities in neighbourhoods of singular points). The structure of the discontinuity curves of T in neighbourhoods of singular points was described in §2. They partition such a neighbourhood into countably many cells, which we have given a natural enumeration, that is, in the order of approximation to the singular point. Again we will consider only those cells that have indices $\geqslant n_1$, while the discontinuity curves bounding the remaining cells will be studied below. Suppose that there is an LUM in the n-th cell. Then (§2) its coefficient of expansion under T is at least const n^d, where $d = 2$ for an LUM in neighbourhoods of singular points of sliding type, while for LUM in neighbourhoods of singular points of infinite horizon type d takes two values: $d = 3/2$ for the subtypes S and SV, and $d = 1$ for the subtype V and the stadium (§2.6). We first consider the simplest cases, when $d = 2$ and $d = 3/2$.

The analysis carried out above allows us to obtain the required estimate for $\Lambda_N(i)$. We assume that the images of the LUM $\gamma^u_{N,i}$ during the first N iterations fall k_1 times in a subdividing strip (with index $\geqslant n_1$), and k_2 times in a cell (also with index $\geqslant n_1$) in neighbourhoods of the singular points. We denote the indices of the corresponding strips by j'_1, ..., j'_{k_1}, and those of the cells by j''_1, ..., j''_{k_2}. Then the following estimate holds:

$$\Lambda_N(i) = \Lambda_0^{\left[\frac{N-k_1-k_2}{m_0}\right]} \cdot \left(\prod_{l=1}^{k_1} j_l'^{\nu}\right) \cdot \left(\prod_{l=1}^{k_2} j_l''^{3/2}\right).$$

We now estimate the influence on the number of components $\gamma^u_{N,i}$ of the discontinuity curves and subdividing segments not taken into consideration in the above. Suppose that an LUM of length D and its future images during m steps do not intersect the subdividing strips with indices $\geqslant n_1$ and cells with the same indices lying in neighbourhoods of the singular points. It then follows from condition B and the construction of the subdividing segments that there is a constant $K_1 > 0$ such that these images of the given LUM intersect at most $K_1 m$ discontinuity curves and subdividing segments, provided that the length D of the original LUM is sufficiently small ($D \leqslant D_0(m, n_1)$).

Here the constant K_1 is larger than the K_0 in condition B, because of the influence of the subdividing segments.

Thus, the inequality (A3.1) can be written in the following form for sufficiently small D ($D \leqslant D_0(m, n_1)$):

$$\text{(A3.2)} \quad p\left(\bigcup_i \gamma^u_{N,i}\right) \leqslant \sum_{k_1, k_2=0}^{N} D \cdot \left(\sum_{n=n_1}^{\infty} n^{-\nu}\right)^{k_1} \times \\ \times \left(\sum_{n=n_1}^{\infty} n^{-3/2}\right)^{k_2} \cdot \Lambda_0^{-\left[\frac{N-k_1-k_2}{m_0}\right]} \cdot (K_1 m)^{[N/m]+1}.$$

In this estimate it is essential that the series $\Sigma n^{-\nu}$ converges, therefore we impose the restriction $\nu > 1$. This is the only place in this paper where we need such a restriction.

It remains to choose, for the given $\nu > 1$, an n_1 so large that $n_1^{-\nu} + (n_1+1)^{-\nu} + \ldots \leqslant \Lambda_0^{-1/m_0}$, and then we obtain the estimate

$$\text{(A3.3)} \qquad p\left(\bigcup_i \gamma^u_{N,i}\right) \leqslant DN^2 \Lambda_0^{-[N/m_0]} (K_1 m)^{[N/m]+1}.$$

Theorem 3.12 can be derived from this estimate without difficulty.

It remains to consider the singular points of subtype V and the case of the stadium (that is, when $d = 1$, see above). In neighbourhoods of these points the coefficient of expansion of an LUM lying in the n-th cell is approximately equal to $C_1 n$ (C_1 = const). However, a single LUM intersects only finitely many cells with indices from N to $C_2 N$ (see Remark 2.2 and Lemma 2.3). Therefore such an LUM gives a contribution to (A3.2) equal to

$$\text{(A3.4)} \qquad \sum_{n=N}^{C_2 N} \frac{1}{C_1 n} \approx \frac{1}{C_1} \ln C_2 = \text{const}.$$

Moreover, cells in a neighbourhood of singular points of subtype V "wander" (Remark 2.2), and hence their total contribution to (A3.3) does not exceed $(C_1^{-1} \ln C_2)^{N\varepsilon}$ for some small $\varepsilon > 0$ (more precisely, $\varepsilon = \varepsilon(n_1)$ will be arbitrarily small for sufficiently large n_1). This contribution does not exert an essential influence on the estimate (A3.3).

Finally, in the case of the stadium the cells under consideration are not "wandering" any more, but we are "rescued" by the precise values of the constants in formula (A3.4) (see Lemma 2.3): $C_1 = 4$ and $C_2 = 9$. Since $4^{-1} \ln 9 < 1$, the general estimate remains true.

References

[1] D.V. Anosov, Geodesic flows on closed Riemannian manifolds of negative curvature, Trudy Mat. Inst. Steklov. **90** (1967). MR **36** # 7157.
= Proc. Steklov Inst. Math. **90** (1967).

[2] ——— and Ya.G. Sinai, Some smooth ergodic systems, Uspekhi Mat. Nauk **22**:5 (1967), 107–172. MR **37** # 370.
= Russian Math. Surveys **22**:5 (1967), 103–167.

[3] S.N. Bernstein, Collected works, Vol. 4, Izdat. Akad. Nauk SSSR, Moscow 1964. MR **30** # 2.

[4] R. Bowen, *Metody simbolicheskoi dinamiki*: *Sb. statei* (Methods of symbolic dynamics: Collection of articles), Mir, Moscow 1979.

[5] L.A. Bunimovich and Ya.G. Sinai, The fundamental theorem of the theory of scattering billiards, Mat. Sb. **90** (1973), 415–431. MR **51** # 3395.
= Math. USSR-Sb. **19** (1973), 407–423.

[6] ———, On billiards close to dispersing, Mat. Sb. **95** (1974), 49–73. MR **49** # 7422.
= Math. USSR-Sb. **23** (1974), 45–67.

[7] ———, Decay of correlations in dynamical systems with chaotic behaviour, Zh. Eksper. Teoret. Fiz. **89** (1985), 1452–1470.
= Soviet Physics JETP **62** (1985), 842–852.

[8] ———, Ya.G. Sinai, and N.I. Chernov, Markov partitions for two-dimensional billiards, Uspekhi Mat. Nauk **45**:3 (1990), 97–134.
= Russian Math. Surveys **45**:3 (1990), 105–152.

[9] I.A. Ibragimov and Yu.V. Linnik, *Nezavisimye i statsionarno svyazannye velichiny*, Nauka, Moscow 1965. MR **34** # 2049.
Translation: Independent and stationary sequences of random variables, Wolters-Noordhoff, Groningen 1971.

[10] I.P. Kornfel'd, Ya.G. Sinai, and S.V. Fomin, *Ergodicheskaya teoria*, Nauka, Moscow 1980. MR **83a**:28017.
Translation: Ergodic theory, Springer-Verlag, New York 1982.

[11] A.N. Livshits, Homology properties of *Y*-systems, Mat. Zametki **10** (1971), 555–564. MR **45** # 2746.
= Math. Notes **10** (1971), 758–763.

[12] Ya.G. Sinai, Markov partitions and *Y*-diffeomorphisms, Funktsional. Anal. i Prilozhen. **2**:1 (1968), 64–89. MR **38** # 1361.
= Functional Anal. Appl. **2** (1968), 61–82.

[13] ———, Dynamical systems with elastic reflections. Ergodic properties of scattering billiards, Uspekhi Mat. Nauk **25**:2 (1970), 141–192. MR **43** # 481.
= Russian Math. Surveys **25**:2 (1970), 137–189.

[14] ———, Gibbs measures in ergodic theory, Uspekhi Mat. Nauk **27**:4 (1972), 21–64. MR **53** # 3265.
= Russian Math. Surveys **27**:4 (1972), 21–69.

[15] ——— and N.I. Chernov, Ergodic properties of certain systems of two-dimensional discs and three-dimensional balls, Uspekhi Mat. Nauk. **42**:3 (1987), 153–174. MR **89c**:58097.
= Russian Math. Surveys **42**:3 (1987), 181–207.

[16] *Sovremennye problemy matematiki*: *Fundamental'nye napravleniya* (Current problems in mathematics: Fundamental directions), Vol. 2, VINITI, Moscow 1985.
Translation: Dynamical systems II, Springer, Heidelberg 1989.

[17] N.I. Chernov, Topological entropy and periodic points of two-dimensional hyperbolic billiards, Funktsional. Anal. i Prilozhen. **25**:1 (1991), 50–57.

[18] J.-P. Bouchaud and P. Le Doussal, Numerical study of a d-dimensional periodic Lorentz gas with universal properties, J. Statist. Phys. **41**:1–2 (1985), 225–248. MR **86k**:82004.

[19] R. Bowen, Equilibrium states and the ergodic theory of Anosov diffeomorphisms, Lecture Notes in Math. **470** (1975). MR **56** # 1364.

[20] ——— and D. Ruelle, The ergodic theory of axiom A flows, Invent. Math. **29** (1975), 181–202. MR **52** # 1786.

[21] L.A. Bunimovich, On the ergodic properties of nowhere dispersing billiards, Comm. Math. Phys. **65** (1979), 295–312. MR **80h**:58037.

[22] ——— and Ya.G. Sinai, Markov partitions for dispersed billiards, Comm. Math. Phys. **77** (1980), 247–280. MR **82e**:58059.

[23] ——— and ———, Statistical properties of Lorentz gas with periodic configuration of scatters, Comm. Math. Phys. **78** (1981), 479–497. MR **82m**:82007.

[24] ——— and ———, Markov partitions for dispersed billiards (Erratum), Comm. Math. Phys. **107** (1986), 357–358. MR **87m**:58090.

[25] ———, A theorem on ergodicity of two-dimensional hyperbolic billiards, Comm. Math. Phys. **130** (1990), 599–621.

[26] G. Casati, G. Comparin, and I. Guarneri, Decay of correlations in certain hyperbolic systems, Phys. Rev. **A26** (1982), 717–719.

[27] G. Gallavotti and D. Ornstein, Billiards and Bernoulli schemes, Comm. Math. Phys. **38** (1974), 83–101. MR **50** # 7480.

[28] Y. Guivarc'h and J. Hardy, Théorèmes limites pour une classe de chaînes de Markov et applications aux difféomorphismes d'Anosov, Ann. Inst. H. Poincaré. Probab. Statist. **24**:1 (1988), 73–98. MR **89m**:60080.

[29] A. Katok and J.M. Strelcyn, Smooth maps with singularities: invariant manifolds, entropy and billiards, Lecture Notes in Math. **1222** (1987).

[30] A. Krámli and D. Szász, The problem of recurrence for Lorentz processes, Comm. Math. Phys. **98** (1985), 539–552. MR **86m**:60063.

[31] ———, N. Simànyi, and D. Szász, A "transversal" fundamental theorem for semi-dispersing billiards, Comm. Math. Phys. **129** (1990), 535–560.

[32] J. Machta, Power law decay of correlations in a billiard problem, J. Statist. Phys. **32** (1983), 555–564. MR **85e**:70010.

[33] D. Ruelle, Thermodynamic formalism, Addison-Wesley, Reading, MA 1978. MR **80g**:82017.

[34] Ya.G. Sinai, Hyperboic billiards, Proc. Internat. Congress of Mathematicians, Kyoto 1990.

[35] D. Smale, Differentiable dynamical systems, Bull. Amer. Math. Soc. **73** (1967), 747–817. MR **37** # 3598. (Erratum: MR **39**, p. 1593.)

[36] F. Vivaldi, G Casati, and I. Guarneri, Origin of long-time tails in strongly chaotic systems, Phys. Review Lett. **51** (1983), 727–730. MR **84k**:58157.

Translated by R.A.M. Hoksbergen

Institute for Oceanology of the USSR Academy of Sciences
Institute of Theoretical Physics of the USSR Academy of Sciences
United Institute of Nuclear Research

Received by the Editors 25 April 1991