Journal of Statistical Physics, Vol. 74, Nos. 1/2, 1994

Statistical Properties of the Periodic Lorentz Gas. Multidimensional Case

N. I. Chernov[1]

Received November 3, 1992; final July 30, 1993

In 1981 Bunimovich and Sinai established the statistical properties of the planar periodic Lorentz gas with finite horizon. Our aim is to extend their theory to the multidimensional Lorentz gas. In that case the Markov partitions of the Bunimovich–Sinai type, the main tool of their theory, are not available. We use a crude approximation to such partitions, which we call Markov sieves. Their construction in many dimensions is essentially different from that in two dimensions; it requires more routine calculations and intricate arguments. We try to avoid technical details and outline the construction of the Markov sieves in mostly qualitative, heuristic terms, hoping to carry out our plan in full detail elsewhere. Modulo that construction, our proofs are conclusive. In the end, we obtain a stretched-exponential bound for the decay of correlations, the central limit theorem, and Donsker's Invariance Principle for multidimensional periodic Lorentz gases with finite horizon.

KEY WORDS: Hyperbolic dynamical systems; periodic Lorentz gas; decay of correlations; Brownian motion.

1. INTRODUCTION

So far only a few models are known in mathematical physics that exhibit nontrivial or even rich chaotic behavior and, at the same time, have been studied with mathematical rigor. One of those models is the planar periodic Lorentz gas for which the theory of billiard-type dynamical systems works. The ergodicity and K-property of that model were proven by Sinai in 1970[16] and its B-property was proven by Gallavotti and Ornstein in 1974.[10] A deep exploration of its statistical properties was done much

[1] Joint Institute for Nuclear Research, Dubna, P.O. Box 79, Moscow, and Department of Mathematics, Princeton University, Princeton, New Jersey 08540. E-mail: chernov@math. Princeton. eku

later, in 1981, by Bunimovich and Sinai[6] and recently their techniques were improved in ref. 7. From the physical point of view, however, the planar gas is not quite realistic—recall that originally H. Lorentz introduced his model in 1905 to describe an electronic gas in metals. As for the Lorentz gas in three and higher dimensions, only its ergodicity and K-property were established by Sinai and the present author in 1987.[17] The present paper is devoted to the statistical properties of the multidimensional Lorentz gas. We obtain here the same properties for that model as were established for the planar gas in refs. 6 and 7.

The periodic Lorentz gas is a dynamical system generated by the free motion of a point particle in the d-dimensional space $\mathbb{R}^d$, $d \geqslant 2$, which collides elastically with fixed scatterers situated in space periodically. As usual, we suppose the speed of the particle to be one and the scatterers to be disjoint and strictly convex with smooth (at least of class C^3) boundaries whose sectional curvature is uniformly bounded away from 0 and ∞.

Assumption A (Finite horizon). The time of free motion between scatterers is uniformly bounded above.

By projecting the particle trajectory down to a suitable d-dimensional torus $\mathbb{T}\mathrm{or}^d$ we can get a dynamical system with a compact phase space denoted by $\mathfrak{M} = Q \times S^{d-1}$, where Q is the torus $\mathbb{T}\mathrm{or}^d$ with a finite number of scatterers removed from it and S^{d-1} is the unit sphere, the space of the velocity vectors. The projection of the motion of our particle down to Q generates a flow $\{\Psi^t\}$ on $\mathfrak{M}$ with a continuous time t. This is a so-called semidispersing billiard system. It preserves the Liouville measure $d\mu = c_\mu \, dq \, dv$, where dq and dv are simply the Lebesgue measures in Q and S^{d-1}, respectively, and c_μ is a normalizing factor.

A discrete-time version of a billiard dynamics is usually constructed by a cross section of the phase space defined as $M = \{x = (q, v) \in \mathfrak{M}: q \in \partial Q, (v, n(q)) \geqslant 0\}$, where $n(q)$ is the inward unit normal vector to ∂Q at q, and $(\cdot, \cdot)$ stands for the scalar product. So, M consists of all the unit vectors attached to the boundary ∂Q and pointing inside Q (outside the scatterers). At each $x \in M$ we denote $\tau(x)$ the first positive time of reflection of the trajectory starting at x, and $Tx = S^{\tau(x)+0}x$ then specifies the first return map $T: M \to M$. The map T preserves the measure $d\nu = c_\nu(v, n(q)) \, dv \, dq$, which is obtained by the projection of the Liouville measure $d\mu$ onto M (c_ν is again a normalizing factor). Both the map T and the flow $\{\Psi^t\}$ are known to be ergodic, mixing, and enjoy the K-property.[17] We will only work with the discrete-time system (T, M, ν).

For the precise statement of our results we introduce the classes of Hölder continuous (HC) and piecewise Hölder continuous (PHC) functions on M. An HC function f satisfies the condition $|f(x)-f(y)| \leqslant C(f)\, \|x-y\|^{\beta}$ for some $\beta > 0$ (the Hölder exponent). A PHC function is a function which is HC on a finite union of subdomains in M separated by a finite number of compact smooth hypersurfaces. For example, $\tau(x)$ and $\tau(T^{-1}x)$ are both PHC functions.

All four theorems formulated below are proven here under Assumption A and one more, technical, Assumption B (see Section 2). The situation when Assumption A fails is discussed briefly in Section 7.

Theorem 1.1 (Decay of correlations). Let $f(x)$ and $g(x)$ be two HC or PHC functions on M. Then

$$|\langle (f \circ T^n) \cdot g \rangle - \langle f \rangle \langle g \rangle| \leqslant c(f, g)\, \alpha^{\sqrt{n}} \tag{1.1}$$

where $c(f, g) > 0$ depends on f, g and $\alpha < 1$ is determined by the configuration of scatterers and the class of HC or PHC function under consideration.

Here and further on $\langle \cdot \rangle$ denotes the expectation with respect to the invariant measure ν.

Theorem 1.2 (Central limit theorem). Again, let $f(x)$ be an HC or a PHC function with $\langle f \rangle = 0$. Then the quantity

$$\sigma^2 = \sum_{n=-\infty}^{\infty} \langle (f \circ T^n) \cdot f \rangle \tag{1.2}$$

is finite and nonnegative. If $\sigma \neq 0$, then the sequence

$$\frac{f(x) + f(Tx) + \cdots + f(T^{n-1}x)}{(\sigma^2 n)^{1/2}} \tag{1.3}$$

converges in distribution to the standard normal law as $n \to \infty$.

Remark (see, e.g., ref. 11). The sum (1.2) equals zero if and only if the function $f(x)$ is a coboundary one, i.e., $f(x) = g(Tx) - g(x)$ a.e. for another function $g \in L_2(M, \nu)$.

Next, we lift the dynamics back up to the space $\mathbb{R}^d$ from Q. The moving particle starts somewhere in the unit cube $[0, 1]^d$ and then travels in $\mathbb{R}^d$ colliding with an infinite array of scatterers. Denote by $\mathbf{q}(t)$ its position in space at time t and by $\mathbf{q}_n$ the point of the nth reflection. The starting position $\mathbf{q}(0)$ (or $\mathbf{q}_0$) is selected randomly according to the probability measure μ (resp., ν).

Theorem 1.3 (Limit distribution of the displacement vector). The vectors

$$\frac{\mathbf{q}(t)-\mathbf{q}(0)}{\sqrt{t}} \quad \text{and} \quad \frac{\mathbf{q}_n-\mathbf{q}_0}{\sqrt{n}} \tag{1.4}$$

both converge in distribution to d-dimensional nondegenerate normal laws with zero means.

The covariance matrices $\mathbf{V}_1$ and $\mathbf{V}_2$ of two normal distributions involved in Theorem 1.3 are known as diffusion matrices. The latter can be expressed by the (discrete-time) Green–Kubo formula

$$\mathbf{V}_2=\frac{1}{2\langle\tau(x)\rangle}\sum_{n=-\infty}^{\infty}\langle(\mathbf{q}_1-\mathbf{q}_0)^T\otimes(\mathbf{q}_{n+1}-\mathbf{q}_n)\rangle \tag{1.5}$$

Here $(\mathbf{q}_1-\mathbf{q}_0)^T$ is a column vector and $(\mathbf{q}_{n+1}-\mathbf{q}_n)$ is a row vector, so that their (tensor) product is a $d\times d$ matrix. The convergence of the infinite series in (1.5) is assured by Theorem 1.1. A continuous-time Green–Kubo formula for V_1 can be also written down, but we are unable to prove it because of a lack of necessary bounds on the decay of correlations for the flow $\{\Psi^t\}$.

The next theorem requires a certain space-time rescaling. For every $s\in[0,1]$ and $t>0$ we denote $\mathbf{q}_t(s)=\mathbf{q}(st)/\sqrt{t}$. The measure μ induces the probability distribution μ_t on the set of all possible trajectories $\mathbf{q}_t(s)$, $0\leqslant s\leqslant 1$, which are then considered as points in the space $C_{[0,1]}(\mathbb{R}^d)$ of continuous vector functions on $[0,1]$.

Theorem 1.4 (Convergence to the Brownian motion). The measure μ_t converges weakly to a Wiener measure.

The planar versions of Theorems 1.1–1.4 were first proven by Bunimovich and Sinai in 1981.[6] Their proofs were based on the Markov partitions of the space M constructed in their previous paper in 1980.[5] Those partitions were used to approximate the dynamical system (T, M, ν) by probabilistic Markov chains with sufficiently strong ergodic and mixing properties. After that certain classical methods from probability theory were applied to derive Theorems 1.1–1.4.

Unfortunately, a direct extension of that approach to the case $d\geqslant 3$ fails due to the absence of suitable Markov partitions. The problems with "nonsmooth boundaries" specific formultidimensional hyperbolic systems were first described by Bowen in 1978.[4] So far they have prevented "explicit" constructions of Markov partitions for multidimensional billiards in the spirit of refs. 3 and 5. In a recent paper by Krüger and Troubetzkoy[14] a somewhat different (in a sense, "implicit") construction of a Markov partition for an abstract nonuniformly hyperbolic system, very close to the Lorentz gas, was presented. That partition is not ready to use for the study of the statistical properties of the system. It is hoped that it can be improved and applied to billiards, but a realization of that program might require hard work.

Our proofs of Theorems 1.1–1.4 bypass Markov partitions. We only use a finite approximation to those partitions, which is, in a sense, very crude. We call it a Markov sieve, since it does not even cover the phase space M. A tiny subset of positive (but small enough) measure is left out. Such an approach has been developed in ref. 7, where the planar versions of Theorems 1.1–1.3 were reproven and also extended to semidispersing billiards and stadia. In a sense, that approach is more straightforward than the one used in the original works(5,6) and its techniques are simpler. After all, the original approach spanned two full articles, which we accomplish in one.

The paper is organized as follows. Section 2 contains the necessary background of the theory of hyperbolic billiards. In Section 3 we extend the notions of homogeneous stable and unstable manifolds and parallelograms introduced in ref. 7 for planar billiards to the multidimensional case. In Section 4 we prove two basic lemmas on the evolution of the homogeneous manifolds. In Section 5 we construct Markov sieves. Section 6 contains the proofs of Theorems 1.1–1.4.

A final remark. We work with a billiard system, and so the complete proofs of our theorems inevitably invoke intricate and very specific techniques of the theory of billiards. However, our principal ideas are very general and no doubt can work for other multidimensional nonuniformly hyperbolic systems, including attractors. In order to make our ideas and arguments easy to understand for the general reader, nonexpert in billiards, we remove the billiard-related technical proofs from the main text and place them in the Appendix. We also try to give emphasis to intuitive, heuristic explanations and descriptions. The last remark especially pertains to Section 5, where the existence of Markov sieves is demonstrated qualitatively rather than proved rigorously.

2. PRELIMINARIES

There have been fairly many papers on billiards published in the past two decades. By now that theory has been sufficiently far developed so that we can effectively use its machinery while avoiding long and tedious calculations typical for early works on this topic. In this section we briefly introduce general well-known facts.

The map T and its inverse T^{-1} are piecewise smooth. Their discontinuities are made up by the trajectories tangent to ∂Q. Denote $S_0 = \partial M = \{x = (q, v) \in M: (v, n(q)) = 0\}$ and $S_i = T^i S_0$ for every integer i. Then S_{-n} is the singularity set for T^n, $n \neq 0$. For $m < n$ denote $S_{m,n} = S_m \cup \cdots \cup S_n$. Obviously, all the powers of T are continuous on a subset $M_c = M \backslash S_{-\infty,\infty}$. For each $m \geqslant 1$ the set $S_{-m,m}$ consists of a finite number of smooth compact hypersurfaces in M with boundaries.

Assumption B (Nonaccumulation property of singularities). The number of smooth components of $S_{-m,m}$ meeting at any point $x \in M$ does not exceed a constant K_0 independent of m.

This assumption holds for generic configurations of scatterers. It is true, for instance, if any trajectory undergoes at most a fixed number of tangent reflections. Assumptions of that kind have been usually made in the literature; see, e.g., refs. 5–7.

The map T is hyperbolic since all the scatterers are strictly convex. The hyperbolicity means that at every point $x \in M_c$ the tangent space $\mathcal{T}_x M$ is decomposed as $E_x^u \oplus E_x^s$, each of $E_x^{u,s}$ being a $(d-1)$-dimensional subspace. This decomposition is DT-invariant, i.e., $DTE_x^{u,s} = E_{Tx}^{u,s}$ at every $x \in M_c$. The space E_x^u corresponds to all the positive Lyapunov exponents, while E_x^s corresponds to the negative ones.

A convenient description of the subspaces $E_x^{u,s}$, $x = (q, v) \in M$, through certain curvature operators has been worked out in nearly all the previous papers on billiards. Take a point $x \in M$ and any $(d-1)$-dimensional submanifolds $\Gamma_1^{u,s}(x)$ in M passing through x and tangent to $E_x^{u,s}$. Each of those manifolds generates a bundle of trajectories in the domain Q outgoing from ∂Q and another bundle incoming to ∂Q. The curvature operator of the orthogonal cross section of the outgoing bunch is denoted by $\mathscr{B}_+^{u,s}(x)$ and that of the incoming one is denoted by $\mathscr{B}_-^{u,s}(x)$. The operators $\mathscr{B}_+^{u,s}(x)$ act in the $(d-1)$-dimensional subspace $J_x \subset \mathbb{R}^d$ orthogonal to the outgoing velocity vector v, and $\mathscr{B}_-^{u,s}(x)$ act in the hyperplane J_{x_-} orthogonal to the incoming velocity vector $v_- = v - 2(v, n(q))\, n(q)$. All those operators are self-conjugate, the $\mathscr{B}_\pm^u(x)$ are positive definite, and the $-\mathscr{B}_\pm^s(x)$ are positive definite, too. In other words, the bundles of trajectories generated by E_x^u are convex (diverging), and those generated by E_x^s are concave (converging).

There are simple equations governing the evolution of the above operators under the action of T. Let $x = (q, v) \in M_c$. First,

$$\mathscr{B}_-^{u,s}(Tx) = \mathscr{B}_+^{u,s}(x) \cdot (I + \tau(x)\, \mathscr{B}_+^{u,s}(x))^{-1} \tag{2.1}$$

Here and further I denotes the identity operator. Second,

$$\mathscr{B}^{u,s}_{+}(x) = \mathscr{B}^{u,s}_{-}(x) + K(x), \quad \text{where} \quad K(x) = 2(v, n(q))\, V^{*}(x)\, K_0(q)\, V(x) \tag{2.2}$$

Here $K_0(q)$ is the curvature operator of the boundary surface ∂Q at q, and V^* and V are two projection operators: V is a projection of the hyperplane J_x onto a hyperplane orthogonal to $n(q)$ along the vector v, and V^* is a projection of the latter back to the former, but now along the vector $n(q)$. The spaces J_x and J_{x_-} can be identified by an isometric projection along the normal vector $n(q)$, and we assumed that identification in (2.1). Combining (2.1) and (2.2), we can express $\mathscr{B}^{u,s}_{\pm}(x)$ as operator-valued continued fractions (see Appendix), but we never use those expressions in the main text.

From (2.1) and (2.2) one can conclude that the eigenvalues of $\mathscr{B}^{u}_{-}(x)$ and $-\mathscr{B}^{s}_{+}(x)$ are uniformly bounded away from 0 and ∞. But two other operators, $\mathscr{B}^{u}_{+}(x)$ and $-\mathscr{B}^{s}_{-}(x)$, may have one very large eigenvalue, roughly proportional to $(v, n(q))^{-1}$, which will cause a lot of trouble in our calculations.

It turns out that natural distances in $E^{u,s}_x$ induced by the Riemannian structures in the phase spaces $\mathfrak{M}$ and M are no good for studying the action of DT, since those spaces are not monotonically expanded or contracted. In order to get monotonicity, we use another coordinate system in $E^{u,s}_x$ induced by the Riemannian structure of the orthogonal cross section to the bundles of trajectories in Q generated by these subspaces (it can be taken either just before the reflection at the point x or after it, the result is the same). In that coordinate system the derivative DT of the map T acts on $E^{u,s}_x$ as

$$DT|_{E^{u,s}_x} = I + \tau(x)\, \mathscr{B}^{u,s}_{+}(x) \tag{2.3}$$

As a result, one gets that E^u_x is expanded by DT in every direction with the factor (rate) uniformly bounded away from 1. The same is true for the contraction of E^s_x. In what follows we refer to these properties of T as simply expansion and contraction. Note that there are presumably no upper bounds on the rates of expansion and contraction, since one eigenvalue of $\mathscr{B}^{u}_{+}(x)$ and $\mathscr{B}^{s}_{-}(Tx)$ may be arbitrarily large.

Some technical remarks. The spaces $E^{u,s}_x$, as well as the operators $\mathscr{B}^{u,s}_{\pm}(x)$, depend continuously on the point $x \in M_c$. The angles between E^u_x and E^s_x in $\mathscr{T}_x M$ (now taken in the Riemannian structure in M) are uniformly bounded away from 0. The angles between the hypersurfaces S_m, $m \geq 0$, and E^s_x, $x \in S_m$, are uniformly bounded away from 0 (this is proven in the Appendix). On the contrary, the angles between S_m, $m \geqslant 0$, and E^u_x, $x \in S_m$, uniformly tend to zero as $m \to \infty$.

The Katok–Strelcyn theory[12] ensures the existence of local stable and unstable manifolds (LUMs and LSMs for brevity) $\gamma^u(x)$ and $\gamma^s(x)$ passing through a.e. point $x \in M_c$. Those LUMs and LSMs are tangent to E^u_x and E^s_x, respectively. Those manifolds have finite sizes (they are compact) due to the discontinuities of both T and T^{-1}. The "unnatural" metric in $E^{u,s}_x$ introduced in (2.3) also induces a special metric in $\gamma^{u,s}_x$ which is used throughout this paper unless otherwise specified. We call it the ρ-metric. It induces a Riemann measure (volume) in LUMs and LSMs, which we also denote by ρ.

An important property of the LUMs and LSMs in billiards is their absolute continuity. It is described in terms of a canonical isomorphism. For any two LUMs γ^u_1 and γ^u_2 sufficiently close to each other, the canonical isomorphism is defined as a map which takes a point $x \in \gamma^u_1$ to the point $\gamma^s(x) \cap \gamma^u_2$ (provided the latter exists). A dual map is defined for any two close LSMs $\gamma^s_{1,2}$. These maps are absolutely continuous with respect to the natural Riemannian measure in $\gamma^{u,s}_1$.[16,13] This property is termed the absolute continuity of the LUMs and LSMs. For two points $x, y \in M$ we denote $[x, y] = \gamma^s(x) \cap \gamma^u(y)$. For two subsets A, $B \subset M$ we denote $[A, B] = \{[x, y]: x \in A, y \in B\}$. For an LUM γ^u and a subset A we denote $\gamma^u_A = \gamma^u \cap A$, and we call the set $\{x \in \gamma^u: \gamma^s(x) \cap A \neq \varnothing\}$ the canonical projection of A onto γ^u.

3. PARALLELOGRAMS

Our principal goal in this section is to extend the notions of homogeneous LUMs, LSMs, and parallelograms elaborated in erfs. 7 and 8 for a planar gas to the multidimensional case. All the necessary proofs are provided in the Appendix.

A parallelogram is a subset $A \subset M$ such that for every pair $x, y \in A$ a point $z = [x, y]$ exists and also belongs in A. Alternatively, $A = [\gamma^u_A(x_0), \gamma^s_A(x_0)]$ for every point $x_0 \in A$. Parallelograms are often called rectangles, but we intentionally follow the terminology of refs. 5–8.

Next, we fix a point $x_0 \in A$, and let $B \subset A$ be a subparallelogram. We denote by Γ^u_B and Γ^s_B the canonical projections of B onto $\gamma^u(x_0)$ and $\gamma^s(x_0)$, respectively. Hence, $B = [\Gamma^u_B, \Gamma^s_B]$. Note that x_0 need not belong in B. If the parallelogram B is an infinitesimal one, its measure can be evaluated as

$$\nu(B) = c_\nu \cdot \det(\mathscr{B}^u_+(x) - \mathscr{B}^s_+(x)) \cdot J^u(x) \cdot J^s(x) \cdot \rho(\Gamma^u_B) \cdot \rho(\Gamma^s_B) \quad (3.1)$$

for $x \in B$. Here the $J^{u,s}(x)$ stand for the Jacobians of the canonical isomorphisms from $\gamma_B^{u,s}(x)$ onto $\Gamma_B^{u,s}$ at x. This is a generalization of a formula established for planar billiards in ref. 7. The proof of (3.1) is given in the Appendix, Section A1.

Now, for an arbitrary subparallelogram $B \in A$ one can easily set up an integral formula

$$\nu(B) = c_\nu \int_{\Gamma_B^u} d\rho(y) \int_{\Gamma_B^s} d\rho(z) \cdot \det(\mathscr{B}_+^u(x) - \mathscr{B}_+^s(x)) \cdot J^u(x) \cdot J^s(x) \quad (3.2)$$

where $y \in \Gamma_B^u$ and $z \in \Gamma_B^s$ are specified by $[y, z] = x$, and both integrals are taken with respect to the ρ-measures in $\gamma^{u,s}(x_0)$.

Next, we fix two numbers $\alpha_0 \in (0, 1)$ and $C_0 > 0$. A parallelogram A is said to be weakly n-homogeneous, $n \geqslant 0$, if

$$|\det(\mathscr{B}_+^u(x) - \mathscr{B}_+^s(x))/\det(\mathscr{B}_+^u(y) - \mathscr{B}_+^s(y)) - 1| \leqslant C_0 \alpha_0^n \quad (3.3)$$

and

$$|J^{u,s}(x) - 1| \leqslant C_0 \alpha_0^n \quad (3.4)$$

for every x, $y \in A$ and every $x_0 \in A$ determining $J^{u,s}(x)$. The main property of a weakly n-homogeneous parallelogram A is that the measure of any subparallelogram $B \subset A$ can be approximated by

$$\nu_a(B) = c_\nu \cdot \det(\mathscr{B}_+^u(x_0) - \mathscr{B}_+^s(x_0)) \cdot \rho(\Gamma_B^u) \cdot \rho(\Gamma_B^s) \quad (3.5)$$

with an exponentially small error:

$$|\nu_a(B)/\nu(B) - 1| \leqslant C_1 \alpha_0^n \quad (3.6)$$

with some $C_1 = C_1(\alpha_0, C_0)$.

An important consequence of (3.5)–(3.6) is the following Markovian approximation formula. A subparallelogram $B \subset A$ is said to be u-inscribed (or s-inscribed) in A (refs. 5, 7) if $\gamma_B^u(x) = \gamma_A^u(x)$ [resp., $\gamma_B^s(x) = \gamma_A^s(x)$] for every $x \in A \cap B$, If B_1 is u-inscribed in A and B_2 is s-inscribed in A, then one has

$$\left| \frac{\nu(B_2/B_1)}{\nu(B_2/A)} - 1 \right| \leqslant C_2 \alpha_0^n \quad (3.7)$$

with some $C_2 = C_2(\alpha_0, C_0)$ provided A is an n-homogeneous parallelogram.

The construction of weakly homogeneous parallelograms is technically very close to that for two-dimensional billiards.[7] First, we fix a $\theta > 1$ and $n_0 \geqslant 1$ and denote by $\mathscr{D}_0$ the infinite union of hypersurfaces in M defined by

$(v, n(q)) = n^{-\theta}$ for all integers $n \geqslant n_0$. These hypersurfaces divide the neighborhood of $S_0 = \partial M$ into an infinite number of thin layers, the closer to S_0 the thinner. An LUM (respectively, an LSM) is said to be homogeneous (we refer to them as HLUM and HLSM for brevity) if it and all its images under T^n for $n < 0$ (resp., for $n > 0$) do not cross the set $\mathscr{D}_0$. An LUM γ^u (an LSM γ^s) is said to be n-homogeneous, $n \geqslant 0$, if $T^n \gamma^u$ (resp., $T^{-n}\gamma^s$) is a homogeneous LUM (LSM). The main reason to introduce homogeneous LUMs and LSMs is that they provide an efficient control on the largest eigenvalue of the operators $K(x)$, $\mathscr{B}^u_+(x)$ and $\mathscr{B}^s_-(x)$. Recall that this eigenvalue is roughly proportional to $(v, n(q))^{-1}$.

A parallelogram A is said to be n-homogeneous, $n \geqslant 0$, if for every point $x \in A$ the sets $\gamma^{u,s}_A(x)$ belong in the same HLUM or, respectively, HLSM. It turns out that n-homogeneous parallelograms are always weakly n-homogeneous with the constants α_0 and C_0 in (3.3)–(3.4) determined by θ and n_0 above. Besides, n-homogeneous LUMs have another important property: for every pair of points x, y in such an LUM

$$|\Lambda^u_k(T^{-k}x)/\Lambda^u_k(T^{-k}y) - 1| \leqslant C_0 \alpha_0^n \tag{3.8}$$

for every $k \geqslant 1$, where $\Lambda^u_k(T^{-k}x)$ stands for the local rate of expansion of the ρ volume in $\gamma^u(T^{-k}x)$ by T^k at the point $T^{-k}x$. A dual estimate to (3.8) holds for LSMs.

The proofs of (3.3), (3.4), and (3.8) are fairly long. They are carried out in the Appendix, Section A2.

We also need some relatively simple properties of homogeneous LUMs and LSMs listed below. Their proofs are outlined in the Appendix, Section A3.

For a.e. point $x \in M$ and every $n \geqslant 0$ there are n-homogeneous LUMs and LSMs passing through x. The maximal smooth components of the HLUMs and HLSMs passing through x are denoted by $\gamma^{0u}(x)$ and $\gamma^{0s}(x)$, respectively.

Remark. Any nonhomogeneous LUM may contain many (possibly, infinitely many) HLUMs inside it. Those HLUMs are separated by the images of the set $\mathscr{D}_0$ intersecting the original LUM. We claim that those images cannot accumulate near any interior point of $\gamma^u(x)$ for a.e. $x \in M$. [That is, they can accumulate only at $\partial\gamma^u(x)$.]

For any $x \in M$ we denote by $r^u(x)$ and $r^s(x)$ the distances of x from $\partial\gamma^{0u}(x)$ and $\partial\gamma^{0s}(x)$, respectively. The following estimate holds for any $\varepsilon > 0$ and some $\beta > 0$:

$$v\{x: r^u(x) < \varepsilon \text{ or } r^s(x) < \varepsilon\} \leqslant \text{const} \cdot \varepsilon^\beta \tag{3.9}$$

The bound (3.9) The bound (3.9) results from another important estimate. Given a subset $A \subset M$, we denote by $U_\varepsilon(A)$ the so-called ε-neighborhood of A in the ρ-metric, i.e., the union of all the stable and unstable manifolds of size $\leqslant \varepsilon$ intersecting A. Then, for some $\beta_1 > 0$, one has

$$\nu(U_\varepsilon(S_{-1,1} \cup \mathscr{D}_0)) \leqslant \varepsilon^{\beta_1} \tag{3.10}$$

4. EVOLUTION LEMMAS

Here we prove two basic lemmas on the evolution of HLUMs and HLSMs in M under the transformation T. We discuss only HLUMs, but dual statements hold for HLSMs in the reverse dynamics.

Let γ^u be an HLUM and ρ_0 denote the normalized ρ volume (ρ measure) in γ^u. The image $T^n\gamma^u$ consists of a finite or countable number of HLUMs called components. At each point $x \in T^n\gamma^u$ we denote by $r_n(x)$ the distance of that point from the boundary of the component of $T^n\gamma^u$ containing x [of course, the distance is measured in the ρ metric defined by (2.3)]. Denote $r_N^{\max}(x) = \max_{0 \leqslant n \leqslant N}\{r_n(T^n x)\}$ for $x \in \gamma^u$ and

$$\bar{r}_n = -\int_{\gamma^u} \ln r_n(T^n x)\, d\rho_0(x)$$

The smaller is γ^u, the larger is the value $\bar{r}_0$ takes. The evolution of a small HLUM under the action of T is determined by two competitive processes. One of them is the expansion which forces the values $r_n(x)$ to grow exponentially fast in n or, equivalently, pushes $\bar{r}_n$ down by a positive amount at each step. The other process is the splitting of $T^n\gamma^u$ into shorter HLUMs when it intersects S_{-1} or $\mathscr{D}_0$, which pushes the mean values $\bar{r}_n$ up again. The next lemma states that the first process is more powerful, so that typical components of $T^n\gamma^u$ will grow in size exponentially fast in n until they reach a certain order of magnitude determined simply by the geometry of the space M.

Lemma 4.1 (Expansion). There are a constant $D > 0$ and a function $\beta(c)$ such that $\beta(c) \to 0$ as $c \to \infty$, and for any HLUM γ^u one has

$$\rho_0\{x \in \gamma^u\colon r_N^{\max}(x) \geqslant D\} \geqslant 1 - \beta(c)$$

with $N = [c\bar{r}_0]$.

In other words, during the first $N \approx \text{const} \cdot \bar{r}_0$ iterates of T the points of γ^u appearing at least once in large (of size $\geqslant D$) components of the images of the HLUM γ^u form a "fat" subset of that HLUM.

The proof of a similar lemma for the planar gas[7] (see also ref. 8) is short and gives even more than stated here. It gives a good estimate for the function $\beta(c)$, which decays exponentially fast as $c \to \infty$. But that proof does not work in the multidimensional case. We outline here another proof, which works in any dimension, but is rather long and gives a much weaker estimate for $\beta(c)$.

Consider a function $F_n(x) = -\ln r_n(T^n x)$ on γ^u. The sequence $\{F_n(x)\}$ equipped with the probability measure ρ_0 can be treated as a discrete-time random process. The expectation with respect to the measure ρ_0 is denoted by $\langle \cdot \rangle_0$, so that $\bar{r}_n = \langle F_n(x) \rangle_0$.

If the discontinuities did not affect the evolution of γ^u, the function $F_n(x)$ would decrease by at least a positive amount $f_0 > 0$ at each iterate of T at each point $x \in \gamma^u$ due to the uniform expansion on HLUMs. In such a case Lemma 4.1 would have followed immediately.

The key point in our arguments is that the splitups of the images of γ^u when they cross S_{-1} or $\mathscr{D}_0$ do not prevent the function $F_n(x)$ from a rapid decrease for typical points $x \in \gamma^u$. First we consider the splitups by S_{-1} alone. Assumption B implies that for each $m \geqslant 1$ there is an $\varepsilon_m > 0$ such that every HLUM of size $\leqslant \varepsilon_m$ intersects no more than K_0 smooth components of $S_{-m,0}$. Thus, T^m cuts any sufficiently small HLUM along no more than K_0 smooth surfaces. That cutting certainly boosts the function $F_n(x)$, but one can bound its "average" increment. The crude idea is that since the number of cutting surfaces is uniformly bounded ($\leqslant K_0$), the total "damage" must be bounded, too.

We now give a precise estimate. Let a component γ_1^u of $T^n\gamma^u$ have a size $\leqslant \varepsilon_m$ and be broken by $S_{-m,0}$ into several subcomponents. For each $x \in \gamma_1^u$ denote by $r_n'(x)$ the ρ distance of x from the boundary of the subcomponent where x belongs. Let $\langle \cdot \rangle_1$ denote the conditional expectation over the HLUM $T^{-n}\gamma_1^u \subset \gamma^u$ equipped with the ρ_0 measure [i.e., $\langle F(x) \rangle_1 = \langle F(x) \rangle_0 / \rho_0(T^{-n}\gamma_1^u)$ for any function $F(x)$ on $T^{-n}\gamma_1^u$].

Sublemma 4.1a. $\langle -\ln r_n'(T^n x) \rangle_1 \leqslant \langle F_n(x) \rangle_1 + f_1$, where the constant $f_1 > 0$ is independent of γ^u or m.

Sublemma 4.1a is proven in the Appendix.

For large m the difference $f_2 = m f_0 - f_1$ is positive. Therefore, the combined effect of the expansion and splitting at m subsequent iterates of T is always a decrease of the average value of the function F_n. Obviously, this is an advantage in our arguments.

We now turn to the splitting of the components caused by the hypersurfaces in $\mathscr{D}_0$. Those can break any component down into an arbitrary large or even infinite number of subcomponents. Hence the above arguments no longer work, and we need a different approach. The situation can be the fact that the rate of expansion of LUMs rapidly grows in the neighborhood of S_0. Roughly speaking, the pieces of a component γ_1^u of $T^n\gamma^u$ broken by $\mathscr{D}_0$ become large enough at the very next step and the effect of expansion outweighs the effect of splitting. Again, let γ_1^u be small enough ($\leqslant \varepsilon_m$) and $\langle \cdot \rangle_1$ denote the conditional expectation over $T^{-n}\gamma_1^u \subset \gamma^u$.

Sublemma 4.1b. $\langle F_{n+1}(x)\rangle_1 \leqslant \langle F_n(x)\rangle_1 - f_3$ with a constant $f_3 > 0$ independent of γ^u.

Sublemma 4.1b is proven in the Appendix, Section A.4.

We now complete the proof of Lemma 4.1. We consider an evolution of the HLUM γ^u under T subject to a special "stopping rule." The reason why we introduce such a rule is that when a component of $T^n\gamma^u$ becomes large enough and contains some points y with $r_n(y) > D$, then such points have already "reached the goal" [recall the definition of $r_N^{\max}(x)$] and we do not need to iterate them any further. Precisely, we define the stopping rule as follows. Whenever a component γ_1^u of $T^n\gamma^u$ contains a nonempty subset $\gamma_{1,0}^u = \{y \in \gamma_1^u : r_n(y) > 2D\}$, we take the D-neighborhood $\gamma_{1,s}^u$ of $\gamma_{1,0}^u$ (in the ρ metric on γ_1^u) and cut $\gamma_{1,s}^u$ out of γ_1^u. The set $\gamma_{1,s}^u$ is stopped ("frozen"), and the remaining part $\gamma_{1,v}^u = \gamma_1^u \backslash \gamma_{1,s}^u$ will then continue evolving under T.

Let us consider more closely a component γ_1^u of $T^n\gamma^u$, for which the above stopping rule applies. Since we have cut that component into two subcomponents, we will redefine the function $r_n(x)$ on the "moving" subcomponent $\gamma_{1,v}^u$. It must be now equal to the ρ distance of $x \in \gamma_{1,v}^u$ from the boundary $\partial\gamma_{1,v}^u$ (instead of $\partial\gamma_1^u$). Denote by $\langle \cdot \rangle_{1,v}$ the conditional expectation over $T^{-n}\gamma_{1,v}^u \subset \gamma^u$. The following relation between the old and new values of the function $r_n(x)$ on the moving subcomponent is analogous to Sublemma 4.1a and proven in Remark A.6 in the Appendix:

$$\langle -\ln r_n^{\text{new}}(T^n x)\rangle_{1,m} \leqslant \langle -\ln r_n^{\text{old}}(T^n x)\rangle_{1,m} + f_4 \frac{P^s}{P^v} \tag{4.1}$$

where $f_4 > 0$ is independent of γ^u, and P^v and P^s stand for the ρ_0 measures of $T^{-n}\gamma_{1,v}^u$ and $T^{-n}\gamma_{1,s}^u$, respectively. The meaning of (4.1) is to bound the increment of $-\ln r_n(x)$ caused by introducing the stopping rule. We do not change $F_n(x)$ on $\gamma_{1,s}^u$.

We apply the stopping rule to each moving component at every step. Thus, we redefine the function $F_n(x)$ at every step. Its increment at the nth step can be estimated due to (4.1):

$$\langle F_n^{\text{new}}(x)\rangle_0 \leqslant \langle F_n^{\text{old}}(x)\rangle_0 + f_4 P_n^s \tag{4.2}$$

where

$$P_n^s = \sum_{\gamma_1^u \subset T^n\gamma^u} \rho_0(T^{-n}\gamma_{1,s}^u)$$

is the ρ_0 measure of the set of points stopped at the nth step, exactly (not before or after it).

The part of $T^n\gamma^u$ that has not been stopped up to the nth step consists of a finite or countable number of HLUMs. Note that the preimage of that "moving" part under T^{-n} has the ρ_0 measure $P_n^v = 1 - \sum_1^n P_i^s$ in the above notations. On the other part of γ^u, which has been stopped before the $(n+1)$th step, we naturally "freeze" the function $F_n(x)$, so that $F_k(x) = F_{k+1}(x) = F_{k+2}(x) = \cdots$ for any point $x \in \gamma^u$ stopped at the kth step.

In order to apply Sublemmas 4.1a and 4.1b to the moving components of $T^n\gamma^u$ we need them to be short enough, i.e., their sizes must be $<\varepsilon_m$. It is not always the case. We apply an additional cutting to ensure that smallness. It is very simple—we just cut "long" components into shorter subcomponents of size $<\varepsilon_m$ along some hyperplanes in M. Those hyperplanes can be selected in a rather arbitrary way, so that the overall increment of the function F_n can be bounded as

$$\langle F_n^{\text{new}}(x)\rangle_0 - \langle F_n^{\text{old}}(x)\rangle_0 \leqslant P_n^v f_5 D/\varepsilon_m \tag{4.3}$$

where f_5 is independent of γ^u or D. This bound is explained in Remark A.7 in the Appendix. We now fix D so small that the RHS of (4.3) will be $< P_n^v f_2/(2m)$. Therefore, the cumulative increment of $\langle F_n(x)\rangle_0$ due to the additional cuttings at any m subsequent steps will be less than $P_n^v f_2/2$.

A combination of Sublemmas 4.1a and 4.1b with the bounds (4.2) and (4.3) gives a bound on the cumulative increment of the mean value of $F_n(x)$ at any m subsequent iterates of T:

$$\langle F_{n+m}(x)\rangle_0 - \langle F_n(x)\rangle_0 \leqslant -P_n^v f_2/2 + (P_{n+1}^s + \cdots + P_{n+m}^s) f_4 \tag{4.4}$$

With a slight abuse of notations, here we denote by $F_n(x)$ the *new* value of this function, after its redefinitions in (4.2) and (4.3) and the above freezing.

The bound (4.4) readily yields for any $n \geqslant 1$

$$\langle F_{mn}(x)\rangle_0 \leqslant \bar{r}_0 - (P_0^v + P_m^v + \cdots P_{m(n-1)}^v) f_2/2 + f_4 \leqslant \bar{r}_0 - nP_{mn}^v f_2/2 + f_4$$

On the other hand, $\langle F_{mn}(x)\rangle_0 \geqslant P_{mn}^v \ln(2D)^{-1}$, because $r_{mn}(x) \leqslant 2D$ on the moving part of $T^{mn}\gamma^u$, whose ρ-measure is P_{mn}^v. Thus,

$$P_{mn}^v \leqslant \frac{\bar{r}_0 + f_4}{nf_2/2 + \ln(2D)^{-1}}$$

and Lemma 4.1 follows.

As a byproduct, we get an explicit formula for the function $\beta(c)$, assuming $\bar{r}_0$ to be large enough, $\beta(c) = 2m/cf_2$. This function decays very slowly as $c \to \infty$, but we conjecture that an exponential bound for P_n^v can be also obtained, as for the planar case in ref. 7. ■

The second evolution lemma pertains to the evolution of sufficiently large HLUMs, i.e., those of size $\geqslant D$ which were obtained in Lemma 4.1. Intuitively, the images of large HLUMs can no longer grow in size due to the compactness of the space M. Instead, the components of the images of such HLUMs in the distant future presumably fill out the space M more or less uniformly. We prove only a weaker version of that conjecture. We call our version the transitivity of HLUMs and HLSMs, as in ref. 7.

To give a precise definition of the transitivity, we have to specify an HLUM γ_1^u to start with and a "place" in the space M which is expected to be filled with the components of $T^n\gamma_1^u$. The HLUM is only supposed to be large enough, i.e., the ρ distance of at least one point x to the boundary $\partial\gamma_1^u$ must be not less than D. To specify an appropriate place in the space M, we need certain geometric notions. Those will be also used later in Section 5.

We fix a large $m \geqslant 1$ and a point $y \in M \backslash S_{-m,m}$. We then fix an arbitrary rectangular coordinate system in the space E_y^u and another one in E_y^s. Together, they form a linear coordinate system in $\mathcal{T}_y M$. Its projection into M by the exponential map determines a coordinate system in a vicinity of $y \in M$. We now take an open cube $V_\varepsilon(y) \subset M$ with sides of length ε parallel to the fixed coordinate axes and centered at y. If ε is small enough, then $V_\varepsilon(y)$ does not intersect $S_{-m,m}$ and looks like a curvilinear parallelepiped in M bounded by some $2(d-1)$ smooth hypersurfaces almost parallel to E_y^u and some other $2(d-1)$ hypersurfaces almost parallel to E_y^s. We call the former the u-sides of $V_\varepsilon(y)$ and the latter the s-sides of it. Certain similar open cubes have been involved in the proofs of the fundamental theorem of the theory of hyperbolic billiards in earlier works.[16,13] Next, we take all the HLUMs inside the cube $V_\varepsilon(y)$ which do not terminate inside $V_\varepsilon(y)$ or on its u-faces, i.e., such HLUMs γ^u that the boundary $\partial(\gamma^u \cap V_\varepsilon(y))$ belongs to the union of s-faces of $V_\varepsilon(y)$. We also take all the HLSMs inside $V_\varepsilon(y)$ which, likewise, do not terminate inside the cube or on its s-faces. Each such a HLUM intersects each HLSM at a single point inside $V_\varepsilon(y)$. The set of the points of intersection is then a homogeneous parallelogram. We denote it by $A_\varepsilon(y)$. We say that such parallelograms are maximal, just as in the planar case.[7,8]

Remark. More generally, given $y \in M \backslash S_{-m,m}$ and two sufficiently small reals $\varepsilon^u, \varepsilon^s > 0$, we can take two open cubes in E_y^u and E_y^s centered at y with sides ε^u and ε^s, respectively. The exponential projection of the direct product of these cubes into M is a curvilinear parallelepiped, which we denote $V_{\varepsilon^u,\varepsilon^s}(y)$. In a similar fashion we define u- and s-faces of $V_{\varepsilon^u,\varepsilon^s}(y)$ and a maximal parallelogram $A_{\varepsilon^u,\varepsilon^s}(y)$ inside it.

For our present purposes we need just one cube $V_0 = V_\varepsilon(y)$ and the corresponding parallelogram $A_0 = A_\varepsilon(y)$ with some $y \in M$ and $\varepsilon > 0$ satisfying the only condition that $\nu(A_0) > 0$. This set A_0 will be fixed throughout the paper. We call it the *meeting place*, since it is a "place" in M where the images and preimages of the elements of the Markov sieve constructed later in Section 5 will meet and intersect one another.

We now turn back to the HLUM γ_1^u. Denote by ρ_1 the normalized ρ measure in γ_1^u. Fix an $n \geqslant 1$. Consider all the components of $T^n\gamma_1^u$ that intersect V_0 and do not terminate inside V_0 or on its u-faces. For each such a component we take its intersection with V_0 and denote the union of all those intersections by $\gamma_{1,n}^u$.

Lemma 4.2 (Transitivity). There are constants $n_0 \geqslant 1$ and $\delta_0 > 0$ determined by V_0 and D alone, such that for every $n \geqslant n_0$,

$$\rho_1(T^{-n}\gamma_{1,n}^u) \geqslant \delta_0 \tag{4.5}$$

In other words, after n_0 iterates of T the portion of the image $T^n\gamma_1^u$ staying inside V_0 (in a "proper way") at any time has a ρ_1 volume bounded away from zero. The constants n_0 and δ_0 do not depend on the HLUM γ_1^u provided it is large enough. The "meeting place" A_0 can be fixed anywhere in the phase space M, thus justifying our term "transitivity."

Sublemma 4.2a. There are a point $x_1 \in \gamma_1^u$ and two reals ε^u, $\varepsilon^s > 0$ such that $\nu(A_{\varepsilon^u,\varepsilon^s}(x_1)) > 0$.

Proof. It is known that for any LUM γ_1^u and for ρ-a.e. point $x \in \gamma_1^u$ there is an LSM $\gamma^s(x)$. Actually, it suffices to have a subset $G \subset \gamma_1^u$ of a positive ρ measure where LSM exist. This statement follows from the so-called fundamental theorem for dispersing billiards.[16,17] The union of the LSMs $\gamma^s(x)$, $x \in G$, has a positive ν measure due to the absolute continuity property of LUMs and LSMs.

We now claim that for a.e. point of G an HLSM $\gamma^{0s}(x)$ also exists. Indeed, if for an $x \in G$, which is an interior point of γ_1^u, there is no HLSM, then either $T^n x \in \mathscr{D}_0$ for some $n \geqslant 0$, or x is a point of accumulation of surfaces separating HLSMs inside the LSM $\gamma^s(x)$. Such "unlucky" points x form a set of zero ρ measure in γ_1^u due to the remark at the end of Section 3.

Thus, $r^s(x) > 0$ at a.e. point $x \in G \subset \gamma_1^u$. Obviously, for some $\delta > 0$ the set $G_\delta = \{x \in \gamma_1^u : r^s(x) > \delta\}$ has a positive ρ measure and is closed. Let x_1 be a density point of the subset G_δ (such that the density of G_δ in small ballshaped neighborhoods of x_1 in γ_1^u is asymptotically one). Besides, we can assume that $\rho(x_1, \partial\gamma_1^u) > \delta$. Now, one can apply all the above arguments to the HLSM $\gamma^{0s}(x_1)$ and easily obtain a maximal parallelogram $A_{\varepsilon^u,\varepsilon^s}(x_1)$ of positive measure. ■

Denote $A_1 = A_{\varepsilon^u,\varepsilon^s}(x_1)$. In virtue of the mixing property of T one has

$$\nu(T^n A_1 \cap A_0) \geqslant \tfrac{1}{2}\nu(A_1)\,\nu(A_0) \tag{4.6}$$

for all $n \geqslant n_1 = n_1(A_1, A_0)$.

The image $T^n A_1$ consists of a finite or countable number of homogeneous parallelograms. Denote by $B_{n,1}$, $B_{n,2}$,... those of them that intersect A_0. If $x' \in B_{n,i} \cap A_0$, then the HLUM $\gamma^{0u}(x')$ is obviously large enough and does not terminate inside V_0 or on its u-faces. The corresponding component of $T^n\gamma_{V_1}^{0u}(T^{-n}x')$ covers the entire $\gamma_{V_0}^{0u}(x')$ unless the point $T^{-n}x'$ is too close to ∂V_1, i.e., unless $\operatorname{dist}(T^{-n}x', \partial V_1) < \alpha^n$ for a certain $\alpha < 1$. This last case is negligible; it pertains to parallelograms whose total relative measure is exponentially small in view of (4.6).

Suppose now that $x' \in B_{n,i} \cap A_0$ and $T^n\gamma_{V_1}^{0u}(T^{-n}x')$ covers the entire HLUM $\gamma_{V_0}^{0u}(x')$. If for any other $x'' \in \gamma^s(x') \cap B_{n,i}$ the image $T^n\gamma_{V_1}^{0u}(T^{-n}x'')$ also covers the entire HLUM $\gamma_{V_0}^{0u}(x'')$, and the set $T^n\gamma_{A_1}^{0u}(T^{-n}x')$ covers the set $\gamma_{A_0}^{0u}(x')$, then the intersection $B_{n,i} \cap A_0$ is "good"—it is a homogeneous parallelogram u-inscribed in A_0 and its preimage $T^{-n}(B_{n,i} \cap A_0)$ is s-inscribed in A_1. This readily follows from the maximality of both A_1 and A_0. We say that such intersections are regular, as in ref. 8. If the intersection $B_{n,i} \cap A_0$ is lacking any of these properties, then it is not large enough, and we say that it is irregular.[8] The latter is the case when either of two following conditions holds: (i) for some point $x'' \in \gamma^s(x') \cap B_{n,i}$ the HLUM $\gamma^{0u}(x'')$ meets either an image $T^k S_{-1}$ or the set $T^k\mathscr{D}_0$ with some k, $1 \leqslant k \leqslant n$, inside V_0, or (ii) the previous condition fails, but for some point $x''' \in \gamma^u(x') \cap A_0$ the HLSM $\gamma^{0s}(T^{-n}x''')$ meets either an image $T^{n-k}S_{-1}$ or the set $T^{n-k}\mathscr{D}_0$ with some k, $1 \leqslant k \leqslant n$, inside V_1. Our method for ruling out such "bad" intersections is different from the one used for the two-dimensional case in ref. 7. A special, nonmaximal, parallelogram A_1 was constructed there, for which bad (irregular) intersections never occurred. Here we allow irregular intersections, but bound the measure of their union by an exponentially small quantity $C\alpha^n$, $\alpha < 1$. Similar ideas were developed earlier in ref. 8. We consider three cases.

Case 1. The value of k is large enough, very close to n. Precisely, let $(1-\delta_1)n \leqslant k \leqslant n$ for some small fixed $\delta_1 > 0$. Then the set $T^{-k}(\gamma^{0u}(x'') \cap B_{n,i})$ lies in the $(c_1\alpha_1^n)$ neighborhood of the union $S_{-1} \cup \mathscr{D}_0$ (in the ρ metric; cf. Section 3) with some $\alpha_1 < 1$. We now estimate the portion of A_1 visiting this tiny neighborhood in the course of $[\delta_1 n]$ iterates of T. The images $TA_1, \ldots, T^{[\delta_1 n]}A_1$ altogether consist of no more than $\Lambda_0^{[\delta_1 n]}$ parallelograms for some $\Lambda_0 > 1$ which is simply determined by the number of the smooth components of the set S_{-1}. Each of those parallelograms intersects the $(c_1 \alpha_1^n)$ neighborhood of $S_{-1} \cup \mathscr{D}_0$ by a set of measure $\leqslant (c_1 \alpha_1^n)^{\beta_1}$, according to the estimate (3.10). Therefore, the total measure of the irregular intersections for the values of k specified above does not exceed const $\cdot \alpha_1^{n\beta_1} \Lambda_0^{\delta_1 n}$. By choosing δ_1 small enough, one can make this bound exponentially small in n.

Case 2. The value of k is very small, $1 \leqslant k \leqslant \delta_2 n$, for some small $\delta_2 > 0$. Similar arguments as in the previous case can now be applied to the parallelogram A_0 and its preimages $T^{-1}A_0, \ldots, T^{-[\delta_2 n]}A_0$. As a result we again get an exponential bound for the total measure of irregular intersections with the current values of k.

Case 3. This case pertains to the remaining, intermediate values of k, which satisfy $\delta_2 n \leqslant k \leqslant (1-\delta_1)n$ with the fixed $\delta_1, \delta_2 > 0$. If $B_{n,i}$ corresponds to such a k, then $T^{-k}B_{n,i}$ is exponentially small in n in every direction due to uniform expansion of LUMs and uniform contraction of LSMs. Therefore, $T^{-k}B_{n,i}$ lies in the $(c_2 \alpha_2^n)$ neighborhood of $S_{-1} \cup \mathscr{D}_0$, where $\alpha_2 < 1$ is determined by the values of δ_1, δ_2. The sets $T^k B_{n,i}$ are disjoint for each fixed k. The estimate (3.10) gives a bound on the measure of the above neighborhood. Summing it over all the related k again yields an exponential bound on the measure of irregular intersections.

Thus, all the irregular intersections are negligible, again in view of (4.6).

Next, in each regular intersection $B_{n,i} \cap A_0$ there is a component of $T^n \gamma_1^u$ which does not terminate inside V_0 or on its s-faces. The inequality (4.5) now formally follows from (4.6).

The last thing we must take care about is the independence of n_0 from the HLUM γ_1^u required by Lemma 4.2. The problem here is that the value of n_1 in (4.6) depends on A_1, which, in turn, may depend on γ_1^u. To solve this problem, we observe that the set H_D of all the "large" HLUMs involved in Lemma 4.2 is compact in the C^0-topology.[7] Indeed, if a sequence of HLUMs converges to a surface in M in C^0-topology, then that surface is easily seen to be an HLUM, too. The parallelogram A_1 in (4.6) can be used not only for a particular HLUM γ_1^u with which it has been constructed, but also for all close HLUMs forming an open neighborhood of that HLUM in H_D. Due to the compactness of H_D a finite collection of such neighborhoods covers H_D, so that a finite number of maximal parallelograms of positive measure can "serve" all the "large" HLUMs. Each of those parallelograms yields its own n_1, and we define n_0 as their maximum. ■

Remark 4.3. As a byproduct of the above proof, we get that for any two maximal parallelograms A_1, A_2 the intersection $T^n A_1 \cap A_2$ for a large $n > 0$ consists mostly of parallelograms that we call regular intersections. Precisely, the measure of the irregular part of $T^n A_1 \cap A_2$ for any $n \geqslant 1$ is bounded above by an exponential function $c\alpha^n$ for some $c > 0$ and $\alpha < 1$ independent of A_1 and A_2.

5. MARKOV SIEVES

Our construction of Markov sieves for the multidimensional Lorentz gas is, in a sense, simpler and cruder than the one elaborated for a planar gas in ref. 7.

We start with an informal description of a Markov sieve—what it looks like and the properties it enjoys. So far it has been used only in refs. 7–9, but it has already proven an effective tool for studying statistical properties of hyperbolic dynamical systems.

A Markov sieve (MS) is a collection of disjoint parallelograms in M. They fill out the entire space M except for a tiny "marginal" set of negligibly small measure. The MS is not a fixed object, unlike the Markov partition. This means that if we are estimating some quantities involved in Theorems 1.1–1.3 for a certain (discrete) time N, then we will construct an MS that works for the given N alone. For different times N we use different MSs, thus working with a one-parameter family of MSs. Furthermore, unlike infinite Markov partitions,[5,6] the elements of an MS for each N are comparable (uniform) in structure and size. Namely, they all are n-homogeneous parallelograms for a certain n (n can be thought of as a second parameter of the MSs, but in our proofs it will be determined by N). The diameters of the elements of an MS are exponentially small in n and the measure of the marginal set is exponentially small in n, too. In other words, for the larger N we take smaller parallelograms, which, however, fill the space M more densely.

We also outline the way the MSs work. Once constructed for a given value of N (and the corresponding n), an MS gives a natural representation of the iterates T^i on M, $1 \leqslant i \leqslant N$, by a stochastic process with a discrete time and with a finite number of states (here the elements of the MS along with the marginal set form the set of states). The first property of that process is sort of a "short memory." It allows us to approximate that process by a finite-state Markov chain. The principal property of the approximating Markov chain is that the convergence to the stationary distribution is fast enough so that it can be observed on the given interval $[0, N]$. In other words, the relaxation to equilibrium essentially needs no more than N iterates of T to occur. In probability theory such properties of Markov chains are called regularity conditions. After that the proofs of

Theorems 1.1–1.3 will be accomplished by invoking standard methods of probability theory.

We now give a formal definition of the MSs. An MS is determined by two large integer parameters N and n, $0<n<N$ [actually, we will consider $n \approx N^{\beta}$ for some $\beta \in (0, 1)$; however, we need a certain freedom in choosing β]. We denote an MS by $\mathfrak{R}_{N,n}$ and its elements by $B_1, \ldots, B_I$, $I=I(n, N)$. We also denote by $B_0 = M \backslash (B_1 \cup \cdots \cup B_I)$ the marginal set. We denote by $\mathfrak{I}$ the set of indices $\{1, \ldots, I\}$, and so $\mathfrak{I}^k$ is the set of k-tuples of indices.

The MS $\mathfrak{R}_{N,n}$ is defined by four conditions. Here and further on $\alpha_1, \alpha_2, \ldots$ stand for various constants in the interval (0, 1) and $c_1, c_2, \ldots$ stand for various positive constants, usually coefficients. The values of α_i and c_i do not depend on the MS parameters N and n.

Condition 1 (Sizes). diam $B_i \leqslant c_1 \alpha_1^n$ for all $i \in \mathfrak{I}$.

Condition 2 (Marginal set). $\nu(B_0) \leqslant c_2 \alpha_2^n$.

Condition 3 (Markovian approximation). For any integers $k>l>1$, $1 \leqslant i_1 < i_2 < \cdots < i_k \leqslant N$, and for any collection $(j_1, \ldots, j_k) \in \mathfrak{I}^k$ one has

$$\begin{aligned} &\nu(T^{i_1} B_{j_1} \cap T^{i_2} B_{j_2} \cap \cdots \cap T^{i_{l-1}} B_{j_{l-1}} / T^{i_l} B_{j_l} \cap \cdots \cap T^{i_k} B_{j_k}) \\ &\quad = \nu(T^{i_1} B_{j_1} \cap \cdots \cap T^{i_{l-1}} B_{j_{l-1}} / T^{i_l} B_{j_l})(1+\Delta) \end{aligned} \tag{5.1}$$

with some $|\Delta| \leqslant c_3 \alpha_3^n$. Here $\nu(B'/B'')$ means the conditional measure, i.e., $\nu(B' \cap B'')/\nu(B'')$.

Condition 4 (Regularity). There are constants g_0, $g_1 > 0$ independent of N and n such that for every $k \geqslant g_0 n$ and for a majority of pairs $(i, j) \in \mathfrak{I}^2$ (see below) one has

$$\nu(T^k B_i \cap B_j) \geqslant g_1 \nu(B_i)\, \nu(B_j) \tag{5.2}$$

The "majority of pairs" means the following. For every i we denote by $R_i(k) \subset \mathfrak{I}$ the collection of the values of j for which (5.2) holds. Then we consider the subset $R(k) \subset \mathfrak{I}$ of integers i such that

$$\sum_{j \in R_i(k)} \nu(B_j) \geqslant 1 - c_4 \alpha_4^n \tag{5.3}$$

Now the "majority of pairs" means exactly that

$$\sum_{i \in R(k)} \nu(B_i) \geqslant 1 - c_5 N \alpha_5^n \tag{5.4}$$

The rest of the present section is devoted to the construction of MSs. It starts in a different manner than that for the planar gas.[7] The latter was based on so-called pre-Markov partitions,[5–8] which partitioned the two-dimensional space M into "nice boxes" ("squares" circumscribed by LUMs and LSMs) whose boundaries enjoyed certain Markovian invariance property; see refs. 6–8 for details. That property was necessary to prevent possible "ugly" mutual intersections of the images of those boxes. Those ugly intersections are illustrated below.

We cannot construct pre-Markov partitions here in the multidimensional case. Moreover, they are unlikely to exist, because the singularity hypersurfaces $S_{-\infty,\infty}$ may cut the boundaries of any "nice" box in a very "ugly" manner and thus rule out any strict invariance property that was valid in the two-dimensional case. But we are able to construct a crude analog of those pre-Markov partitions, and it works well enough. After all, the MS only provides an approximation to the dynamics, and so its construction and properties need not be absolutely rigid, they are flexible by their nature.

Our construction is based on the ideas of Sinai[15] and Bowen.[3] First we fix a large integer $m \geqslant 1$. The set $M \backslash S_{-m,m}$ consist of a finite number of domains with piecewise smooth boundary. We fill those domains with some boxes $W_i = V_\varepsilon(x_i)$, $1 \leqslant i \leqslant I_1$ (see Section 4), where $\varepsilon = e^{-n}$. The boxes should cover the entire space M except for a vicinity of the set $S_{-m,m}$. Precisely, the boxes cover $M \backslash U_{c'_6 \alpha_6^n}(S_{-m,m})$ and do not intersect $U_{c''_6 \alpha_6^n}(S_{-m,m})$ with some $c'_6 > c''_6 > 0$ [here $U_\varepsilon(S)$ stands for the ε-neighborhood of a set S].

Locally, the centers of boxes constitute a nearly regular $(2d-2)$-dimensional lattice with a variable spacing ranging from 0.8ε to 0.9ε. The boxes are aligned so that neighboring boxes have nearly parallel faces. Thus each box intersects $3^{2(d-1)} - 1$ neighboring boxes and the sizes of overlappings are of order ε. Similar covers have been used in refs. 13 and 17 and we omit details. Provided m is large enough, such boxes exist, because we allow the constants c'_6, c''_6, and α_6 to depend on m, but not on n.

Next, we adjust the boundaries of the boxes to prevent certain "ugly" intersections of the images of those boxes under T^m and T^{-m} with other boxes. Consider the image $T^m W_i$ of a box W_i. It looks like a "pancale"—largely extended in all the u-directions but extremely thin in s-directions. In Fig. 1 three possible intersections of $T^m W_i$ with another box W_j are shown. We say that the first two are "ugly" and our next goal is to rule them out. First note that the last intersection shown in Fig. 1c is a typical one, but the two ugly ones still occur too often and can spoil further estimates if we do not adjust the boundaries of the boxes.

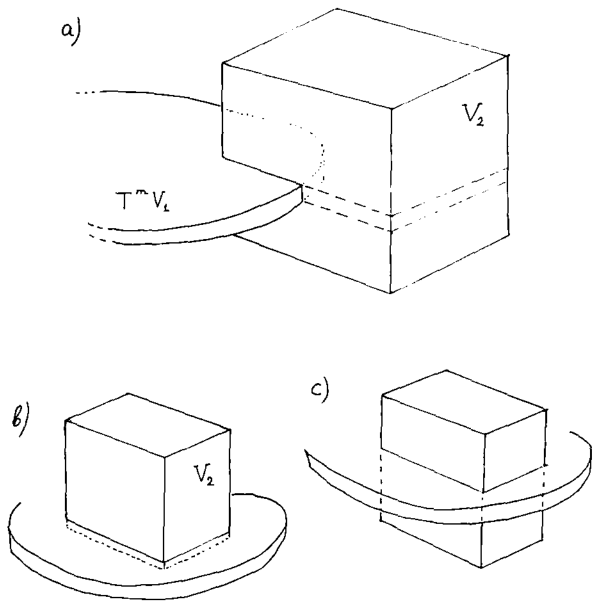

Fig. 1. Possible intersections of the "pancake" $T^m W_i$ and the box W_j: (a, b) "ugly" ones; (c) a typical one.

The idea of the adjustment is borrowed from ref. 15. In the case shown in Fig. 1a we expand the box W_i in the u-directions, so that the image $T^m W_i^{(1)}$ of the expanded box $W_i^{(1)}$ will intersect the box W_j very neatly—it will spread exactly up to the s-faces of the latter; see the dashed lines in Fig. 1a. The case shown in Fig. 1b is symmetric to the preceding one because the intersection $T^{-m} W_j \cap W_i$ looks exactly like that shown in Fig. 1a. It is treated in a similar fashion—the box W_j is expanded in the s-direction, so that it will reach the boundaries of $T^m W_i$.

After all the adjustments we get slightly larger boxes W_i, $1 \leqslant i \leqslant I_1$, of a less regular shape, however. For our convenience we may assume that they still have a cylindrical form, i.e., they have a direct product structure in the local coordinate system induced by $E^{u,s}_{x_i}$; see Section 4. Furthermore, we can modify the boundary of the new box so that it will consist of a finite number of hypercubes, which we continue calling u- and s-faces. In that case we have to sacrifice "the neatness" of the intersections of the images of the new boxes with the old ones, now allowing tiny slots between their s-faces, but possible losses are easily seen to be negligible. Actually, the losses (in measure) are of the same order of magnitude as the slots between the original boxes, i.e., exponentially small in n.

Our adjustments have reduced the "ugliness" of the intersections, but have not eliminated it completely. Indeed, the boundary of the box W_j in Fig. 1a has been adjusted, too, so that some new gaps appear between the s-faces of $W_j^{(1)}$ and those of $T^m W_i^{(1)}$. Hence, we need a new adjustment consisting in further expansion of the new boxes in such a way that the images of their new boundaries fit the boundaries of the boxes obtained at the previous step. This second adjustment results in somewhat larger boxes $W_i^{(2)}$, $1 \leqslant i \leqslant I_1$. Additionally, we modify new boxes again, so that they will have cylindrical structure and their boundaries will consist of a finite number u- and s-faces, as we did above. After performing those adjustments (followed by modifications) k times, we get some boxes $W_i^{(k)}$, $1 \leqslant i \leqslant I_1$, of a rather irregular shape. Clearly, the increments of the boxes decay exponentially fast in k, and when m is large, the total increments are small enough compared to the sizes of the original boxes (say, they are less than $\varepsilon/100$ if m is large enough).

We stop this iterational procedure at a finite step $k = n$, avoiding some serious troubles with nonsmooth boundaries of the limiting "boxes" as $k \to \infty$; see refs. 3 and 15 for more detail. Our boxes $W_i^{(n)}$ are still cylindrical sets and still have a finite number of u- and s-faces. Their boundaries are not completely adjusted—there are some tiny slots [of width smaller than $\varepsilon\delta^n = (\delta e^{-1})^n$, $\delta < 1$] between the s-faces of images of boxes and other boxes as described above. However, possible losses are negligible.

Next, we borrow another idea from refs. 3 and 15 to define smaller but disjoint boxes here. First we observe that each box intersects exactly $3^{2(d-1)}$ neighboring boxes (except for the outermost boxes, adjacent to an uncovered vicinity of $S_{-m,m}$). For each pair V_i and V_j of intersecting boxes we partition each of them into four smaller ones. To specify that partition we observe that the spaces $E^{u,s}_{x_i}$ are almost parallel to $E^{u,s}_{x_j}$, because x_i and x_j are close. Thus, we can assume that the box V_j has a direct product structure in the coordinate system $E^{u,s}_{x_i}$. Say, let $V_i = E^u_i \times E^s_i$ and $V_j = E^u_j \times E^s_j$, where $E^{u,s}_{i,j}$ are subregions in the spaces $E^{u,s}_{x_i}$. We call those regions the basic regions of the boxes V_i and V_j, respectively. Actually, only the box V_i has a direct product structure in this coordinate system, and so our assumption requires a slight perturbation of V_j, but the relative error gained from that is exponentially small in n, and so it is negligible. We now partition the box V_i into four ones:

$$V^1_{i,j} = (E^u_i \cap E^u_j) \times (E^s_i \cap E^s_j)$$

$$V^2_{i,j} = (E^u_i \backslash E^u_j) \times (E^s_i \cap E^s_j)$$

$$V^3_{i,j} = (E^u_i \cap E^u_j) \times (E^s_i \backslash E^s_j)$$

$$V^4_{i,j} = (E^u_i \backslash E^u_j) \times (E^s_i \backslash E^s_j)$$

All the four boxes $V^k_{i,j}$, $1 \leqslant k \leqslant 4$, have direct product structures in our coordinate system and a finite number of u- and s-faces, and they are disjoint.

Note that we do not disturb the (nearly) invariance property of the boundaries gained above, since each face of the old boxes is extended no farther than by the distance ε and at least by the distance 0.05ε. (That observation may, however, fail for boxes V whose images $T^m V$ and $T^{-m} V$ are close to $S_{-m,m}$, but their total measure is exponentially small.)

We now consider the collection of all "parts" $V^k_{i,j}$, $1 \leqslant k \leqslant 4$, $1 \leqslant i \neq j \leqslant I_1$. Since the boxes $V^1_{i,j}$ and $V^1_{j,i}$ almost coincide, we discard either one of them for every pair i, j. Given a point $x \in M$, we define a box $V(x) = \bigcap \{V^k_{i,j} : x \in V^k_{i,j}\}$. These boxes are disjoint and cover the whole M except for a tiny set of exponentially small measure. They enjoy the same (approximate) invariance property of the boundaries as the old boxes did.

Some of our new boxes might be anomalously short. We will simply remove such boxes. To be specific, we retain a box $V = E^u_V \times E^s_V$, where E^u_V and E^s_V are its basic regions, if each of those regions contains at least one point such that the ρ distance of it from the boundary of that region is not less than $c_7 \alpha_7^n$. If $\alpha_7 < e^{-1}$ is small enough, then the total measure of the removed boxes is exponentially small in n, and so the losses are negligible. We fix an $\alpha_7 \in (\delta e^{-1}, e^{-1})$.

Thus far we have worked with T^m instead of T in order to ease the control on the possible increments of the boxes during the adjustments. But now we have to turn back to T. Denote by $\mathscr{V}$ the set of boxes constructed above and take $\mathscr{W} = \mathscr{V} \vee T\mathscr{V} \vee \cdots \vee T^{m-1}\mathscr{V}$. This is the collection of all the mutual intersections of those boxes and their images under $T, \ldots, T^{m-1}$. For our convenience we also split the disconnected elements of this collection into their connected components. As a result we get a collection of smaller boxes, which have a direct product structure and a finite number of u- and s-faces. Again, we remove all the anomalously short boxes from $\mathscr{W}$ by the same rule as above (with the same value of α_7). The new system of boxes enjoys the invariance property of the boundary under the action of both T and T^{-1} (we always mean an approximate invariance, allowing gaps of width $\leqslant c_7 \alpha_7^n$, which are exponentially smaller than the sizes of our boxes). It is our crude analogue of a pre-Markov partition.

We now start the construction of the Markov sieve (MS). For each box $V \in \mathscr{W}$ we denote by E^u_V and E^s_V the basic regions and then take their $(2c_7\alpha_7^n)$ neighborhoods $\hat{E}^u_V$ and $\hat{E}^s_V$ [in the corresponding $(d-1)$-planes]. The box $\hat{V} = \hat{E}^u_V \times \hat{E}^s_V$ has a direct product structure and contains V. We call $\partial \hat{E}^s_V \times \hat{E}^u_V$ (respectively, $\partial \hat{E}^u_V \times \hat{E}^s_V$) the u-boundary (s-boundary) of $\hat{V}$. Inside each box $V \in \mathscr{W}$ we take all the HLUMs spreading up to the s-boundary of $\hat{V}$. This means that the HLUMs do not terminate inside $\hat{V}$

or on its u-boundary. Likewise, we take all the HLSMs spreading up to the u-boundary of $\hat{V}$. The points of intersections of those HLUMs and HLSMs inside V make up a homogeneous parallelogram $A = A(V) \subset V$. The measure of the set not covered by the parallelograms, $\nu(\bigcup_{\mathscr{W}} (V \backslash A(V))$, is exponentially small in n by virtue of (3.9). Moreover, we can retain only "very dense" parallelograms, i.e., such that

$$\rho(E_V^{u,s} \cap A(V)) \geqslant (1 - c_8 \alpha_8^n)\, \rho(E_V^{u,s}) \tag{5.5}$$

where $E_V^{u,s}$ are the basic regions of the box V (recall that $V = E_V^u \times E_V^s$). If the constant $\alpha_8 < 1$ is sufficiently close to 1, then the total measure of the parallelograms lacking the property (5.5) for either E_V^u or E_V^s is exponentially small in n , and so it is negligible. Denote the resulting collection of parallelograms by $\mathscr{A}$. The approximate invariance property of the boundaries of the boxes $V \in \mathscr{W}$ under $T^{\pm 1}$ and our definition of boxes $\hat{V}$ readily imply the regularity of the intersections of $T^{\pm 1} A_1 \cap A_2$ for all A_1, $A_2 \in \mathscr{A}$, i.e., those intersections are either empty or regular.

We now take $\mathscr{A}_n = T^{-n}\mathscr{A} \vee \cdots \vee \mathscr{A} \vee T\mathscr{A} \vee \cdots \vee T^n \mathscr{A}$. This is the collection of the mutual intersections of the parallelograms in $\mathscr{A}$ with their images under T^i for all i, $|i| \leqslant n$. The elements $A \in \mathscr{A}_n$ are n-homogeneous parallelograms covering the entire space M except for a tiny marginal set $A^{(0)}$ whose measure is less than $c_9 \alpha_9^n$.

The intersections $T^{\pm 1} A_1 \cap A_2$ for any two parallelograms $A_1, A_2 \in \mathscr{A}_n$ are either empty or regular, because the same property holds for the elements of $\mathscr{A}$. But this is no longer valid for the intersections $T^{\pm k} A_1 \cap A_2$ with $k \geqslant 2$, since some intermediate images $T^{\pm i} A_1$, $1 \leqslant i \leqslant k-1$, may intersect the marginal set $A^{(0)}$ where we lose control on their further evolution. Nonetheless, the regularity of the intersections $T^k A_1 \cap A_2$, $|k| \geqslant 2$, would be important at least for $|k| \leqslant N$ to ensure the condition 3 of the MS. To achieve this regularity, we reduce the parallelograms in $\mathscr{A}_n$ by removing from every $A \in \mathscr{A}_n$ all the points $x \in A$ whose images under T^i, $|i| \leqslant N$, visit the marginal set $A^{(0)}$ at least once. The measure of the set of points removed from all the parallelograms $A \in \mathscr{A}$ does not exceed $2Nc_9\alpha_9^n$. If n is of order N^β with some fixed $\beta > 0$, then the total losses are still exponentially small in n. It is easy to verify, by induction in k, that the remaining subset of each $A \in \mathscr{A}_n$ is still a parallelogram, and the intersections $T^k A' \cap A''$, $A', A'' \in \mathscr{A}_n$, are regular for $|k| \leqslant N$. Moreover, we can discard all the parallelograms which, after the reduction, lack the "density" property (5.5), since their total measure would not exceed $Nc_{10}\alpha_{10}^n$.

The result of the above construction is the desired Markov sieve $\mathfrak{R}_{n,N}$. It obviously satisfies the conditions 1 and 2. The third condition readily follows from the n-homogeneity of the parallelograms in $\mathfrak{R}_{n,N}$, the regularity of the intersections between their images and the relation (3.7).

The verification of the last condition of the MS requires additional considerations. They are based on the two evolution lemmas and involve the "meeting place" A_0 from Section 4. We describe those considerations briefly, since they almost repeat the corresponding arguments for the planar gas.[7]

Take any parallelogram $A \in \mathscr{A}_n$. We first claim that its image $A' = T^n A$ is of size $\geqslant ce^{-n}$ in every u direction. This means that for any point $x \in A'$ and any line $l \in E^u_x$ the projection of A' onto l has the ρ diameter $\geqslant ce^{-n}$. Indeed, for any point $y \in A'$ the set $\gamma^u_{A'}(y)$ lies in an HLUM $\gamma^u_{V'}(y)$, where V' is a box from $\mathscr{W}$. Therefore, this HLUM is of ρ size $\geqslant ce^{-n}$ in every u direction with a value of c determined by m. Besides, $\gamma^u_{A'}$ is a "very dense" subset of that HLUM, precisely,

$$\rho(\gamma^u_{A'}(y)) \geqslant (1 - c_8 \alpha^n_8)\, \rho(\gamma^u_{V'}(y)) \tag{5.6}$$

Therefore, the projection of the set $\gamma^u_{A'}(y)$ onto any u-directed line in the box V' has ρ size $\geqslant ce^{-n}$.

The further images $T^{n+i}A$, for $i \geqslant 1$, may suffer from the discontinuities of T, and so they generally consist of a finite number of homogeneous parallelograms. We call them p-components to distinguish them from the components of the HLUMs discussed earlier in Section 4.

Lemma 4.1 applies to the HLUM $\gamma^u_{V'}(y)$ and says that during the evolution of that HLUM under T^i, $1 \leqslant i \leqslant c_{11} n$, a majority of its points appear at least once in large components of $T^i \gamma^u_{V'}(y)$, i.e., in components whose ρ size in every u direction is $\geqslant D$. Each of these components "carries" on it a p-component of $T^i A'$. The density of the p-components of $T^i A'$ on the corresponding components of $T^i \gamma^u_{V'}(y)$ does not change significantly under the action of T^i by virtue of (3.8). Hence this density on most of the components is still high, as high as was specified by (5.6), with, say, somewhat larger value of the constant c_8.

The evolution Lemma 4.2 now applies to each of the above large components of $T^i \gamma^u_{V'}(y)$, $1 \leqslant i \leqslant c_{11} n$, and says that for every $i \geqslant c_{11} n + n_0$ a certain portion (of relative ρ measure $\geqslant c_{12}$) of the HLUM $\gamma^u_{V'}(y)$ is transformed by T^i into a set of components lying inside the box V_0 and not terminating inside it or on its u-faces. The box V_0 was defined in Section 4. The p-components of $T^i A'$ carried on the components of the images of $\gamma^u_{V'}$ that lie inside V_0 are still dense enough, as specified by (5.6), and so their total measure is at least $\frac{1}{2} c_{12} \nu(A')$.

All the p-components of $T^i A' = T^{n+i} A$ carried on the components of $T^i \gamma^u_{V'}$ inside V_0 certainly intersect the meeting place A_0, but here we need more. We need to work with only the p-components intersecting A_0 regularly, as defined in Section 4. Remark 4.3 provides a good bound for the measure of all the irregular intersections—this bound is exponentially small in n. At first sight, this seems to be not enough, since the measure of each A' is of the same order of magnitude, i.e., it is exponentially small in n. However, we can combine all the irregular intersections of p-components of $T^{n+i}A$ for all $A \in \mathscr{A}_n$ with a fixed i and estimate their total measure. Since those irregular intersections are all disjoint for any fixed i, Remark 4.3 applies to their union and gives the same bound for the total measure of those irregular intersections: $c\alpha^n$ with $c>0$ and $\alpha<1$ independent of n and i. From this bound, one readily gets the following. For each $i \geqslant n + c_{11}n + n_0$ there are some "bad" elements of $\mathscr{A}_n$ whose images under T^i have not enough p-components intersecting A_0 regularly. However, the total measure of those "bad" elements of $\mathscr{A}_n$ is exponentially small in n. All the other elements $A \in \mathscr{A}_n$ have enough p-components of $T^{n+i}A$ intersecting A_0 regularly. Precisely, the total measure of those p-components is at least $\frac{1}{2}c_{12}\nu(A)$. For a given i, we denote the set of "good" elements of $\mathscr{A}_n$ by $\mathscr{A}_n^+(i)$.

We now take another parallelogram $B \in \mathscr{A}_n$. Its preimages $T^{-j}B$, $j \geqslant 1$, have the properties symmetric to the ones of the images of A. As a result, for every $j \geqslant n + c_{11}n + n_0$ there is a dominant collection of elements of $B \in \mathscr{A}_n$ whose images under T^{-j} have enough p-components intersecting A_0 regularly [for each B their total measure is at least $\frac{1}{4}c_{12}\nu(B)$]. For a given j, we denote that dominant collection of $B \in \mathscr{A}_n$ by $\mathscr{A}_n^-(j)$. Note that the regularity of intersections of $T^{-j}B$ with A_0 now means that they are s-inscribed in A_0 and their images under T^j are u-inscribed in B.

We now fix a $k \geqslant 2(n + c_{11}n + n_0)$ and a decomposition $k = i + j$ with i, j such that $\min\{i, j\} \geqslant n + c_{11}n + n_0$. For every pair of parallelograms $A \in \mathscr{A}_n^+(i)$ and $B \in \mathscr{A}_n^-(j)$ we can take advantage of the approximative formulas (3.5)–(3.6) to estimate from below the total measure of the intersections of T^iA with $T^{-j}B$ within the "meeting place" A_0. As a result, one easily gets

$$\nu(T^iA \cap T^{-j}B) \geqslant g_1' \, \nu(A) \, \nu(B) \tag{5.7}$$

with some $g_1' > 0$ determined by the values of c_{12} and C_1 from (3.6).

Next, we consider the elements $\hat{A}$ and $\hat{B}$ of the MS $\mathfrak{R}_{n,N}$ contained in A and B, respectively. In virtue of (5.5), which holds for both pairs A, B and $\hat{A}$, $\hat{B}$, the elements $\hat{A}$ and $\hat{B}$ are dense enough in A and B. Therefore, (5.7) implies

$$\nu(T^{i+j}\hat{A} \cap \hat{B}) \geqslant \tfrac{1}{2} g_1' \, \nu(\hat{A}) \, \nu(\hat{B})$$

for sufficiently large n. Thus, we have proven (5.2) for a majority of pairs $\hat{A}$, $\hat{B} \in \mathfrak{R}_{n,N}$. The estimates (5.3) and (5.4) readily result from our definitions of $\mathscr{A}_n^+(i)$ and $\mathscr{A}_n^-(j)$.

A family of Markov sieves $\mathfrak{R}_{n,N}$ characterized by Conditions 1–4 is now constructed. The following theorem is a direct consequence of Conditions 1–4:

Theorem 5.1 (Relaxation to the equilibrium distribution). For any integers $k \geqslant l > 1$ and $1 \leqslant i_1 < i_2 < \cdots < i_k \leqslant N$ there is a subset $R_* = R_*(i_1, \ldots, i_k) \subset \mathfrak{Z}^{k-l+1}$ of $(k-l+1)$-tuples of indices such that:

(i) If $(j_l, \ldots, j_k) \in R_*$, then

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

(ii) One has

$$\sum_{(j_l, \ldots, j_k) \in R_*}^{l} \nu(T^{i_l}V_{j_l} \cap \cdots \cap T^{i_k}V_{j_k}) \geqslant 1 - \Delta$$

where $\Delta = \max\{c_{13}N^2\alpha_{11}^n, (1-g_1)^{[L/2]}\}$ with $L = [(i_l - i_{l-1})/(g_0 n)]$.

Theorem 5.1 is proven in ref. 7. Note that it is still true if one reverses "the time," i.e., for $N \geqslant i_1 > \cdots > i_k \geqslant 1$. The meaning of Theorem 5.1 is that the conditional distributions relax to equilibrium exponentially fast in the parameter $|i_l - i_{l-1}|$ (which represents the "interval" between the "future" and the "past"), at least as long as it is less than const $\cdot\, n^2$.

6. THE PROOFS OF THEOREMS 1.1–1.4

Theorem 1.1 follows from Theorem 5.1 immediately; see refs. 7 and 8.

The proof of Theorem 1.2 is based on Theorem 5.1 and purely probabilistic arguments, which do not make use of any specific feature of the underlying dynamical system. The proof may be found in ref. 7. Its probabilistic part has been borrowed from ref. 11. We emphasize that MSs, along with Theorem 5.1, are thus a rather universal tool—they automatically ensure the statistical properties described in Theorems 1.1 and 1.2.

The proof of Theorem 1.3 is a combination of some general probabilistic arguments with one special geometric property of the dynamical system in question. That property is only required to ensure the nondegeneracy of the matrices $\mathbf{V}_1$ and $\mathbf{V}_2$. We discuss that last property here and refer for the probabilistic part of the proof of Theorem 1.3 to refs. 6 and 7.

We verify only the nondegeneracy of $\mathbf{V}_2$, since that of $\mathbf{V}_1$ will then follow immediately. Suppose that $\mathbf{V}_2$ is degenerate, i.e., det $\mathbf{V}_2 = 0$. In virtue of our remark to Theorem 1.2, a certain linear combination $(\mathbf{a}, \mathbf{q}_1 - \mathbf{q}_0)$ is then a coboundary function on M. Here $\mathbf{a} = (a_1,..., a_d)$ is a constant vector, and $(\cdot, \cdot)$ again stands for the scalar product in $\mathbb{R}^d$.

All we need now is a periodic point $x \in M$ for the map T with a period $n_0 \geqslant 1$ (i.e., $T^{n_0}x = x$) such that

$$(\mathbf{a}, (\mathbf{q}_{n_0} - \mathbf{q}_0)) \neq 0 \tag{6.1}$$

at the point x. Indeed, the function $(\mathbf{a}, (\mathbf{q}_n - \mathbf{q}_0))$ is the sum of iterates of a coboundary function $(\mathbf{a}, \mathbf{q}_1 - \mathbf{q}_0)$, and so it stays bounded in distribution as n grows. On the other hand, that function grows linearly in n at a particular periodic point. From its continuity, it also takes values proportional to n in a vicinity of the periodic point, which depends on n. If the above coboundary function were bounded, we would get an obvious contradiction. For coboundary functions from $L_2(M, \nu)$ some extra reasonings are required to get a contradiction. They are provided in ref. 7. We only prove here that a periodic point with the property (6.1) exists.

Fix an integer $N \geqslant 1$ and take N^d copies of the torus $\mathbb{T}\mathrm{or}^d$ such that they make up a big cube of size $N \times N \times \cdots \times N$ in the space $\mathbb{R}^d$. This cube can be also considered as a "big" torus $\mathbb{T}\mathrm{or}^d_N$ by imposing periodic boundary conditions. Projecting the trajectory of the moving particle from $\mathbb{R}^d$ down to that big torus, we get a new billiard dynamics in a region Q_N made up by N^d copies of Q (as of "bricks"). We denote by $\mathfrak{M}_N = Q_N \times S^{d-1}$ the phase space, by $\{\Psi^t_N\}$ the phase flow, and by T_N the billiard ball map of that big system. The dynamics $\{\Psi^t_N\}$ obviously commutes with space shifts of the big torus along the sides of the original torus $\mathbb{T}\mathrm{or}^d$. (Those shifts generate a finite Abelian group isomorphic to a direct product of d identical N-element cyclic groups.)

The billiard dynamics in Q_N certainly meets all the conditions of Theorem 1.1, and so we can again consider a parallelogram A_0 defined in Section 4. Since its closure $\bar{A}_0$ is also a parallelogram, we can assume that A_0 is closed. Let $n_1,..., n_d$ be integers such that $0 \leqslant n_i < N$ for every i. Consider the translation of $\mathbb{T}\mathrm{or}^d_N$ generated by n_1 shifts along the first side of the original torus $\mathbb{T}\mathrm{or}^d$, then n_2 shifts along the second side, etc. The resulting translation of $\mathbb{T}\mathrm{or}^d_N$ generates a transformation of the phase space $\mathfrak{M}_N$, which commutes with the dynamics $\{\Psi^t_N\}$. Denote by A' the image of A_0 under that transformation. It has the same form and measure as A_0, it is just located in another part of the phase space. Due to the results of Section 4, the intersection $T^n_N A_0 \cap A'$ for large n consists mostly of the regular intersections of the p-components of $T^n_N A_0$ with A'. Let U be one of those p-components of $T^n_N A_0$. By projecting the dynamics $\{\Psi^t_N\}$ down to the original torus we get a periodic point $x = \cdots \cap T^{-n}U \cap U \cap T^n U \cap T^{2n} U \cap \cdots$ that belongs in M, whose period is n. On the other hand, $\mathbf{q}_n - \mathbf{q}_0$ at the point x is an integer vector with the components $(m_1 N + n_1, m_2 N + n_2, ..., m_d N + n_d)$ with some integers $m_1, ..., m_d$.

Lemma 6.1. For any nonzero real-valued vector $\mathbf{a} = (a_1, ..., a_d)$ one can find integers N and $n_1, ..., n_d$ such that the sum

$$a_1(m_1 N + n_1) + a_2(m_2 N + n_2) + \cdots + a_d(m_d N + n_d) \tag{6.2}$$

does not vanish for any integers $m_1, ..., m_d$.

Actually, it is easier to prove that the expression (6.2) never takes values of the form Nm with any integer m. This can be done by induction in d. The proof is elementary and we leave it to the reader.

The existence of a periodic point $x \in M$ with the property (6.1) is now established. Hence the proof of Theorem 1.3 is accomplished.

Finally, the proof of Theorem 1.4 is based solely on Theorem 1.3 and certain general measure-theoretic arguments. It is provided in ref. 6.

7. CONCLUDING REMARKS

Here we discuss possible extensions of Theorems 1.1–1.4.

First of all, one might wish to relax Assumption A. A periodic Lorentz gas that does not satisfy it is said to have no horizon. Such a gas looks, in a sense, more realistic than the one with finite horizon. For two-dimensional case Theorems 1.1 and 1.2 are still true without Assumption A, as was shown in ref. 7, but Theorem 1.3 fails in that case. The reason is that while the mean value $\langle \tau(x) \rangle$ is still finite, the second moment $\langle \tau^2(x) \rangle$ is infinite, and so the sum (1.5) contains an infinite term at $n = 0$. P. Bleher has conjectured that in this case the displacement $\mathbf{q}_n - \mathbf{q}_0$ in (1.4) should be rescaled by $(n \ln n)^{1/2}$ instead of $\sqrt{n}$. He supported his conjecture with a detailed calculation of the mean displacement $\langle \mathbf{q}_n - \mathbf{q}_0 \rangle$ in ref. 1, but this is not a conclusive proof yet.

We conjecture that Theorems 1.1 and 1.2 are valid for the multidimensional Lorentz gas without horizon as well. The proofs in the planar case[7] involved an explicit description of the so-called "cells" in the space M—the subregions where the map T is continuous and the function $\tau(x)$ takes very large values. It is hoped that the structure of cells for the multidimensional gas can be described in a like fashion. It is easy to verify that the mean $\langle \tau(x) \rangle$ is finite but $\langle \tau^2(x) \rangle$ is infinite, just as for the planar Lorentz gas without horizon. Hence, Theorem 1.3 fails, but a certain rescaling of (1.4) might be possible.

Theorem 1.1 provides only an upper bound for the decay of correlations. The actual rate of the decay is still unknown. Certain old numerical experiments (see, e.g., ref. 2) have revealed a stretched-exponential decay of the type

$$|\langle (f \circ T^n) \cdot g \rangle - \langle f \rangle \langle g \rangle| \asymp \alpha^{n^\gamma}$$

with $\gamma < 1$ (besides, $\gamma \to 1$ as the dimension d grows). On the other hand, certain recent numerical researches showed an exponential decay even for planar hyperbolic billiards (see references given in ref. 8). There is also a rigorous result[8] supporting the conjecture of an exponential rate of the decay of correlations. On the contrary, nothing is known about the decay of correlations for the continuous-time dynamics $\{\Psi^t\}$. Apparently, without the finite horizon Assumption A, the correlations for that flow decay very slowly, most likely as slowly as algebraically.

Finally, we discuss small perturbations of the Lorentz gas dynamics with finite horizon by small constant external fields (electric, magnetic, or combined ones). A perturbation of that type for the planar gas was studied recently in ref. 9. This perturbation destroyed the invariant measure ν for the map T, but strong hyperbolic properties of T persisted. Despite the absence of an absolutely continuous invariant measure, the machinery of the Markov sieves worked. It was used to construct a (singular) invariant measure, to establish its ergodicity, and to estimate the rate of the decay of correlations. The resulting invariant measure was the so-called Sinai–Bowen–Ruelle (SBR) measure, which was singular on M but absolutely continuous on the unstable manifolds. The analysis performed in ref. 9 led to mathematical proofs of certain classic electrodynamic equations, in particular, Ohm's law and the Einstein relation. The results of ref. 9 can be extended straightforwardly to any dimension $d \geqslant 3$ by using the Markov sieves constructed in the present paper. There is a hope that other hydrodynamic equations can be derived in this way.

APPENDIX

The Appendix consists of four sections, in which we give support to the claims made in Sections 3 and 4 whose proofs involve specific billiard techniques. Let us first introduce some conventions. We will denote by α various constants between 0 and 1 whose exact values are not relevant. We also denote by c and a various positive constants.

A1. Here we derive the formula (3.1). We consider a generic point $x = (q, v) \in M$ and define several coexisting measures in the tangent space

$\mathcal{T}_x M$ (all of them will be simply scalar multipliers of the Lebesgue measure). This space is a direct product of the tangent spaces $\mathcal{T}_q(\partial Q)$ and $\mathcal{T}_v S^{d-1}$, and so the first natural measure is the Lebesgue measure $dm_x = dm_x^q \, dm_x^v$, i.e., the product of the Lebesgue measures in $\mathcal{T}_q(\partial Q)$ and $\mathcal{T}_v S^{d-1}$. Another measure is $d\nu_x = c_\nu (v, n(q)) \, dm_x$. We then consider the tangent spaces $\mathcal{T}_x \gamma^{u,s}(x)$ to the local manifolds $\gamma^{u,s}(x)$ at the point x. Their natural projections onto $\mathcal{T}_q(\partial Q)$ along $\mathcal{T}_v S^{d-1}$ are one-to-one linear maps, and so the measure dm_x^q induces on these tangent spaces measures $dm_x^{u,s}$. We also have the measures $d\rho_x^{u,s}$ in the spaces $\mathcal{T}_x \gamma^{u,s}(x)$ induced by the ρ measures in $\gamma^{u,s}(x)$ defined in Section 2. Evidently, $d\rho_x^{u,s} = (v, n(q)) \, dm_x^{u,s}$.

Since $\gamma^u(x)$ and $\gamma^s(x)$ are transversal at x, the product $dm_x^u \, dm_x^s$ is a measure in $\mathcal{T}_x M$. We now derive the relation between this measure and dm_x. To this end, we pick a basis $(e_1, ..., e_{d-1})$ of orthogonal unit vectors in $\mathcal{T}_q(\partial Q)$ such that the vector e_1 is a linear combination of v and $n(q)$. Denote by J_x a $(d-1)$-dimensional subspace in $\mathbb{R}^d$ orthogonal to the velocity vector v of the point x. It is naturally identified with $\mathcal{T}_v S^{d-1}$. Denote by e_1' the unit vector in J_x $(= \mathcal{T}_v S^{d-1})$ that is proportional to the projection of e_1 onto J_x along v. The collection $(e_1', e_2, ..., e_{d-1})$ is a basis in J_x, and then in $\mathcal{T}_v S^{d-1}$. Therefore,

$$\mathbf{E} = ((e_1, 0), (e_2, 0), ..., (e_{d-1}, 0), (0, e_1'), (0, e_2), ..., (0, e_{d-1}))$$

is a basis in $\mathcal{T}_x M$, where a pair (u_1, u_2) of vectors $u_1 \in \mathcal{T}_q(\partial Q)$ and $u_2 \in \mathcal{T}_v S^{d-1}$ means a vector in $\mathcal{T}_x M$ in the usual sense. We now project the vectors $e_1, ..., e_{d-1} \in \mathcal{T}_q(\partial Q)$ onto the subspaces $\mathcal{T}_x \gamma^{u,s}(x) \subset \mathcal{T}_x M$ along $\mathcal{T}_v S^{d-1}$. The projection of e_i is $(e_i, \mathcal{B}_+^{u,s}(x) \, e_i)$ for $i \geqslant 2$, and the projection of e_1 is $(e_1, (v, n(q)) \, \mathcal{B}_+^{u,s}(x) \, e_1')$. One can easily write down the coordinates of all these $2d-2$ vectors in the basis $\mathbf{E}$ as in the form of a $(2d-2) \times (2d-2)$ matrix $\mathbf{J}$ and compute its determinant. The result is

$$|\det \mathbf{J}| = (v, n(q)) \det(\mathcal{B}_+^u(x) - \mathcal{B}_+^s(x))$$

Therefore,

$$\begin{aligned} dm_x &= (v, n(q)) \det(\mathcal{B}_+^u(x) - \mathcal{B}_+^s(x)) \, dm_x^u \, dm_x^s \\ &= (v, n(q))^{-1} \det(\mathcal{B}_+^u(x) - \mathcal{B}_+^s(x)) \, d\rho_x^u \, d\rho_x^s \end{aligned}$$

As a result, we get

$$d\nu_x = c_\nu \det(\mathcal{B}_+^u(x) - \mathcal{B}_+^s(x)) \, d\rho_x^u \, d\rho_x^s$$

and thus complete the proof of the expression (3.1).

A2. We now turn to the inequalities (3.3), (3.4), and (3.8). We derive them together by studying how smoothly the operators $\mathscr{B}^{u,s}_{\pm}(x)$ depend on x in homogeneous parallelograms. First, we write down an operator-valued continuous fraction formula for $\mathscr{B}^{s}_{+}(x)$:

$$\mathscr{B}^{s}_{+}(x) = -\cfrac{I}{\tau_0 I + \cfrac{I}{K(Tx) + \cfrac{I}{\tau_1 I + \cfrac{I}{K(T^2 x) + \cdots}}}} \tag{A.1}$$

where $\tau_i = \tau(T^i x)$ and I stands for the identity operator. A ratio A/B always means AB^{-1}. All the operators in (A.1) act in the $(d-1)$-dimensional space J_x orthogonal to the velocity vector v. Recall that we identified the linear spaces J_x and $J_{T^{-1}x}$ in Section 2. In the same way we now identify the spaces $J_{T^i x}$ and $J_{T^{i+1}x}$ for all $i \in \mathbb{Z}$ by isometric projections along the normal vectors to ∂Q at the points $T^i x$. Thus, all the spaces $J_{T^i x}$, $i \in \mathbb{Z}$, are identified and the formula (A.1) is justified. The operator $\mathscr{B}^{u}_{-}(x)$ is expressed by an operator-valued continued fraction constructed by tracking the negative semitrajectory of x (in this case, however, one has a positive sign in front of the fraction).

We now consider two points $x = (q_x, v_x)$ and $y = (q_y, v_y)$ in an n-homogeneous parallelogram A. Due to uniform contraction and expansion, diam $A \leqslant \text{const} \cdot \alpha^n$, where the diameter is measured in the Riemann metric in M. (Actually, the contraction property has been established for the ρ metric on LUMs and LSMs, but one can easily verify that $\text{dist}(x, y) \leqslant \text{const} \cdot [\rho(x, y)]^{1/2}$ for any x and y in one LUM or LSM.)

Lemma A.1. If two points x and y belong in the same n-homogeneous parallelogram, then

$$\left| \frac{(v_x, n(q_x)) - (v_y, n(q_y))}{(v_x, n(q_x))} \right| \leqslant \text{const} \cdot \alpha^n$$

Proof. We first observe that

$$|(v_x, n(q_x)) - (v_y, n(q_y))| \leqslant \text{const} \cdot \alpha^n \tag{A.2}$$

If x and y are not too close to ∂M, say, if $|(v_x, n(q_x))| \geqslant n_0^{-\theta} = \text{const}$, then Lemma A.1 follows from (A.2). If x and y lie between two close hypersurfaces in $\mathscr{D}_0$, say, $(v, n(q)) = k^{-\theta}$ and $(v, n(q)) = (k+1)^{-\theta}$, $k \geqslant n_0$, then

$$|(v_x, n(q_x)) - (v_y, n(q_y))| \leqslant 2\theta k^{-(\theta+1)} \leqslant 4\theta (v_x, n(q_x))^{(\theta+1)/\theta}$$

Therefore,

$$\left|\frac{(v_x, n(q_x)) - (v_y, n(q_y))}{(v_x, n(q_x))}\right| \leqslant \text{const} \cdot \alpha^{n/(\theta+1)}$$

and Lemma A.1 follows. ■

We now define a special linear transformation T_{xy} on $\mathbb{R}^d$ that takes x to y and maps J_x onto J_y. First we shift $\mathbb{R}^d$ by the vector $q_x\vec{q}_y$, so that q_x will go to q_y. Then we apply a rotation through the angle between v_x and v_y, so that the image of v_x will coincide with v_y. Thus, J_x is mapped onto J_y. In addition, in case $(v_x, n(q_x))$ is small, say, less than $n_0^{-\theta}$, then we apply a rotation of $\mathbb{R}^d$ about the vector v_y so that the principal eigenvector of $K(x)$ (that with the largest eigenvalue) will go to its counterpart for the operator $K(y)$. So, the map T_{xy} is a composition of a shift and one or two rotations. Clearly, T_{xy} is an isometry. Since $K(x)$ depends on x smoothly, the map T_{xy} differs from the identity operator by less than const $\cdot\,\alpha^n$ (in the Euclidean norm).

Denote $\tilde{\mathscr{B}}^{u,s}_{\pm}(y) = T_{xy}^{-1} \circ \mathscr{B}^{u,s}_{\pm}(y) \circ T_{xy}$; these are operators in J_x, just like $\mathscr{B}^{u,s}_{\pm}(x)$. Now, it is a straightforward estimation that

$$\begin{aligned}
&|\det(\mathscr{B}^u_+(x) - \mathscr{B}^s_+(x)) - \det(\mathscr{B}^u_+(y) - \mathscr{B}^s_+(y))| \\
&\quad \leqslant \text{const} \cdot [(v_x, n(q_x))^{-1} \operatorname{dist}(x, y) \\
&\qquad + (v_x, n(q_x))^{-1} - (v_y, n(q_y))^{-1} \\
&\qquad + \|\mathscr{B}^u_-(x) - \tilde{\mathscr{B}}^u_-(y)\| + \|\mathscr{B}^s_+(x) - \tilde{\mathscr{B}}^s_+(y)\|]
\end{aligned} \tag{A.3}$$

The two last terms are the most difficult to bound from above.

Lemma A.2. If two points x and y belong in one n-homogeneous parallelogram A, then

$$\|\mathscr{B}^u_-(x) - \tilde{\mathscr{B}}^u_-(y)\| \leqslant \text{const} \cdot \alpha^n \tag{A.4}$$

and

$$\|\mathscr{B}^s_+(x) - \tilde{\mathscr{B}}^s_+(y)\| \leqslant \text{const} \cdot \alpha^n \tag{A.5}$$

One might think of (A.4) and (A.5) as the Hölder continuity of the operators $\mathscr{B}^u_-$ and $\mathscr{B}^s_+$ on homogeneous parallelograms.

Before proving Lemma A.2, we observe that it, along with Lemma A.1 and the bound (A.3), completes the proof of the inequality (3.3). Thus, we now have to prove (A.4), (A.5), (3.4), and (3.8). The four are closely related and we derive them together.

For any point $x \in M_c$ we denote by

$$\lambda^u(x) = \det(I + \tau(x)\,\mathscr{B}^u_+(x)) \quad \text{and} \quad \lambda^s(x) = [\det(I - \tau(x)\,\mathscr{B}^s_-(Tx))]^{-1}$$

the local one-step rates of expansion of the ρ volume in $\gamma^u(x)$ and, respectively, of contraction of the ρ volume in $\gamma^s(x)$ [cf. (2.3)]. Then

$$\Lambda^{u,s}_k(x) = \lambda^{u,s}(x)\,\lambda^{u,s}(Tx)\cdots\lambda^{u,s}(T^{k-1}x)$$

are the corresponding k-step rates of expansion/contraction. It is well known (see, e.g., refs. 7, 9, 10) that if x and y belong to one LSM (LUM), then the Jacobian of the canonical isomorphism from $\gamma^u(x)$ to $\gamma^u(y)$ [respectively, from $\gamma^s(x)$ to $\gamma^s(y)$] is

$$J^u(x, y) = \lim_{k \to \infty} \frac{\Lambda^u_k(x)}{\Lambda^u_k(y)} \quad \left(\text{resp., } J^s(x, y) = \lim_{k \to \infty} \frac{\Lambda^s_k(T^{-k}y)}{\Lambda^s_k(T^{-k}x)}\right) \tag{A.6}$$

There is a helpful duality in billiard systems. The reverse dynamics $\{\Phi^{-t}\}$ is also a billiard system in the same configuration space Q. So, many statements have their dual forms obtained by just reversing the dynamics. For example, the bounds (A.4) and (A.5) are dual to each other. It always suffices to prove either one of two dual statements.

By taking the natural logarithm of both sides of (A.6), one can easily reduce the estimates (3.4) and (3.8) to two inequalities

$$\left|\ln \frac{\lambda^u(x)}{\lambda^u(y)}\right| \leqslant C'_0 \alpha_0^n \tag{A.7}$$

for any points x and y in the same n-homogeneous LSM and

$$\left|\ln \frac{\lambda^s(x)}{\lambda^s(y)}\right| \leqslant C'_0 \alpha_0^n \tag{A.8}$$

for any points x and y in the same n-homogeneous LUM, with another positive constant C'_0.

Note that (A.7) and (A.8) are *not* dual. Nonetheless, the bound (A.8) does have a dual form: that is the bound (A.7) for any x and y in the same n-homogeneous LUM. In other words, it suffices to prove (A.7) for any x and y in one n-homogeneous parallelogram.

We now derive (A.7) assuming x and y belong in one n-homogeneous parallelogram A. We again invoke the transformation T_{xy}. It is a straightforward calculation that

$$|\det(I+\tau(x)\,\mathscr{B}^u_+(x))-\det(I+\tau(y)\,\mathscr{B}^u_+(y))|$$
$$\leqslant \text{const}\cdot[|\tau(x)-\tau(y)|\,(v_x, n(q_x))^{-1}$$
$$+(v_x, n(q_x))^{-1}-(v_y, n(q_y))^{-1}+(v_x, n(q_x))^{-1}\cdot \text{dist}(x, y)$$
$$+\|\mathscr{B}^u_-(x)-\tilde{\mathscr{B}}^u_-(y)\|] \tag{A.9}$$

It is easily seen that

$$|\tau(x)-\tau(y)|\leqslant \text{dist}(x, y)+\text{dist}(Tx, Ty)\leqslant \text{const}\cdot\alpha^n$$

A combination of the bound (A.9) and Lemmas A.1 and A.2 then gives (A.7) for any $x, y\in A$.

As a result, all our considerations boil down to two inequalities in Lemma A.2, only one of which has to be proven, due to the duality principle. Unfortunately, neither is easy to prove. Our proof is based on a somewhat cumbersome decomposition of the operator-valued continued fraction (A.1).

We will prove the bound (A.5). First we denote $\tilde{\tau}_i=\tau(T^iy)$ and $\tilde{K}(T^iy)=T^{-1}_{xy}K(T^iy)\,T_{xy}$ for $i\geqslant 0$. Then we write down an expression

$$\tilde{\mathscr{B}}^s_+(y)=(\tilde{\tau}_0 I+(\tilde{K}(Ty)+(\tilde{\tau}_1 I+(\tilde{K}(T^2y)+\cdots)^{-1})^{-1})^{-1})^{-1}$$

similar to (A.1). Our further arguments will be based on the decomposition

$$\begin{aligned}\mathscr{B}^s_+(x)-\tilde{\mathscr{B}}^s_+(y)&=\mathscr{B}^s_+(x)\{[\tilde{\mathscr{B}}^s_+(y)]^{-1}-[\mathscr{B}^s_+(x)]^{-1}\}\,\tilde{\mathscr{B}}^s_+(y)\\&=-\mathscr{B}^s_+(x)(\tau_0-\tilde{\tau}_0)\,\tilde{\mathscr{B}}^s_+(y)\\&\quad+D^s_+(x)[K(Tx)-\tilde{K}(Ty)]\,\tilde{D}^s_+(x)\\&\quad+D^s_+(x)[\mathscr{B}^s_+(Tx)-\tilde{\mathscr{B}}^s_+(Ty)]\,\tilde{D}^s_+(y)\end{aligned} \tag{A.10}$$

where

$$D^s_+(x)=\{I+\tau(x)[\mathscr{B}^s_-(Tx)]^{-1}\}^{-1}$$
$$\tilde{D}^s_+(y)=\{I+\tilde{\tau}(y)[T^{-1}_{xy}\mathscr{B}^s_-(Ty)\,T_{xy}]^{-1}\}^{-1}$$

Iterating the decomposition (A.10) k times yields

$$\begin{aligned}&\mathscr{B}^s_+(x)-\tilde{\mathscr{B}}^s_+(y)\\&\quad=-\sum_{i=1}^{k}E^{(i-1)}_+(x)\cdot\mathscr{B}^s_+(T^{i-1}x)\cdot(\tau_{i-1}-\tilde{\tau}_{i-1})\cdot\tilde{\mathscr{B}}^s_+(T^{i-1}y)\cdot\tilde{E}^{(i-1)}_+\\&\qquad+\sum_{i=1}^{k}E^{(i)}_+\cdot[K(T^ix)-\tilde{K}(T^iy)]\cdot\tilde{E}^{(i)}_+\\&\qquad+E^{(k)}_+\cdot[\mathscr{B}^s_+(T^kx)-\tilde{\mathscr{B}}^s_+(T^ky)]\cdot\tilde{E}^{(k)}_+\end{aligned} \tag{A.11}$$

where $E_+^{(i)} = D_+^s(x) \cdots D_+^s(T^{i-1}x)$ and $\tilde{E}_+^{(i)} = \tilde{D}_+^s(T^{i-1}y) \cdots \tilde{D}_+^s(y)$. We stop the decomposition (A.11) at $k = [n/2]$.

In order to prove the bound (A.5), we observe the following:

(i) The operators $D_+^s(T^ix)$ and $\tilde{D}_+^s(T^iy)$ are contractions: $\|D_+^s(T^ix)\| \leqslant \alpha$ and $\|\tilde{D}_+^s(T^iy)\| \leqslant \alpha$ uniformly in x, y and $i \geqslant 0$.

(ii) $\operatorname{dist}(T^ix, T^iy) \leqslant \text{const} \cdot \alpha^{n-i}$, and so $|\tau_i - \tilde{\tau}_i| \leqslant \text{const} \cdot \alpha^{n-i}$.

(iii) $\|\mathscr{B}_+^s(T^kx) - \tilde{\mathscr{B}}_+^s(T^ky)\|$ is uniformly bounded in x, y, and k.

(iv) If, given an $i \leqslant k$, the points $T^ix = (q_i, v_i)$ and T^iy are not too close to ∂M, say, if $(v_i, n(q_i)) \geqslant n_0^{-\theta} = \text{const}$, then

$$\|K(T^ix) - \tilde{K}(T^iy)\| \leqslant \text{const} \cdot \operatorname{dist}(T^ix, T^iy) \leqslant \text{const} \cdot \alpha^{n-i}$$

(v) The last case to be considered is the one when $T^ix = (q_i, v_i)$ and T^iy belong in one thin layer between two close hypersurfaces of $\mathscr{D}_0$ for some $i \leqslant k$. In this case the operators $K(T^ix)$ and $\tilde{K}(T^iy)$ have one large eigenvalue each. [For example, if the scatterer at which the reflection at the point T^ix occurs is a sphere of radius r, the largest eigenvalue of $K(T^ix)$ is $2r^{-1}(v_i, n(q_i))^{-1}$.] The difference $K(T^ix) - \tilde{K}(T^iy)$ might also have a large norm, however. In this case we will estimate the norm of a composition

$$F_i = D_+^s(T^{i-1}x) \cdot (K(T^ix) - \tilde{K}(T^iy)) \cdot \tilde{D}_+^s(T^iy)$$

All our arguments will be based on elementary geometric considerations and we only outline the main steps. Denote by χ_i and $\tilde{\chi}_i$ the largest eigenvalues of $K(T^ix)$ and $\tilde{K}(T^iy)$, respectively, and e_i and $\tilde{e}_i$ the corresponding unit eigenvectors. First we notice that the points T^jx and T^jy are ε_n-close for all $j \leqslant n/2$ with some $\varepsilon_n = \text{const} \cdot \alpha^n$. Thus, $\|e_i - \tilde{e}_i\| \leqslant \text{const} \cdot \alpha^n$ and, due to Lemma A.1, $|\chi_i - \tilde{\chi}_i| \leqslant \text{const} \cdot \alpha^n \chi_i$. We now take an arbitrary unit vector $w \in J_x$ and consider $w_1 = \tilde{D}_+^s(T^{i-1}y)\, w$. Since $\tilde{\chi}_i$ is large, it is easy to verify that $|(w_1, \tilde{e}_i)| \leqslant \text{const} \cdot \tilde{\chi}_i^{-1}$. Besides, $|(w_1, e_i)| \leqslant \text{const} \cdot (\chi_i^{-1} + \varepsilon_n)$. We then consider a vector $w_2 = (K(T^ix) - \tilde{K}(T^iy))\, w_1$. Its projection onto e_i has a length $\leqslant \text{const} \cdot \chi_i \varepsilon_n$, and its projection onto the orthogonal complement to e_i in J_x has a length $\leqslant \text{const} \cdot \varepsilon_n^a$ for some $a > 0$. Finally, the vector $w_3 = D_+^s(T^{i-1}x)\, w_2$ has a length $\leqslant \text{const} \cdot \varepsilon_n^a$. Thus, we conclude that $\|F_i\| \leqslant \text{const} \cdot \varepsilon_n^a$.

Applying all the five observations (i)–(v) to the decomposition (A.11) results in the bound (A.5). Lemma A.2 is now proven.

A3. We proceed by supporting the relatively simple properties of HLUMs and HLSMs mentioned at the end of Section 3. First we verify the inequality (3.10). Every HLSM of ρ size $\leqslant \varepsilon$ that intersects the set $S_{-1,1} \cup \mathscr{D}_0$ wholly lies in a $(\text{const} \cdot \sqrt{\varepsilon})$ neighborhood of $S_{-1,1} \cup \mathscr{D}_0$ (taken in the Riemannian metric on M). The ν-measure of this neighborhood can be estimated directly, and one gets the bound (3.10).

Our next step is almost compiled from ref. 7, Appendix 2. We pick a $p_0 > 0$ and set $\varepsilon_n = p_0 \lambda_0^n$ for all $n \geq 0$, where $\lambda_0 < 1$ is an upper bound on the one-step rate of contraction of HLSMs in the ρ metric. We then estimate

$$\sum_{n=0}^{\infty} \nu(U_{\varepsilon_n}(S_{-1,1} \cup \mathscr{D}_0)) \leqslant \text{const} \cdot p_0^{\beta} \tag{A.12}$$

We now claim that, given a point $x \in M$ such that $T^n x \notin U_{\varepsilon_n}(S_{-1,1} \cup \mathscr{D}_0)$ for all $n \geqslant 0$, an HLSM $\gamma^s(x)$ exists and the ρ distance from x to its boundary $\partial\gamma^s(x)$ is bounded below by p_0. The proof of this claim is well known; see, e.g., similar statements in refs. 12 and 16. Thus, the bound (A.12) implies (3.9).

Furthermore, let an LSM γ^s contain an infinite number of HLSMs and the surfaces separating them accumulate at an interior point of γ^s. Then the images $T^n\gamma^s$, $n \geqslant 1$, intersect $\mathscr{D}_0$ for an infinite sequence of values of n. Due to the inequality (A.12) and the Borel–Cantelli lemma such points x form a set of zero measure.

A4. In this section of the Appendix we support the claims made in the proof of Lemma 4.1. We will work here in the "full" phase space $\mathfrak{M}$ instead of M. The flow $\{\Phi^t\}$ is hyperbolic, and so at a.e. point $x = (q, v) \in \mathfrak{M}$ there is a $(d-1)$- dimensional local unstable manifold $\Gamma^u(x)$ for the flow. Its natural projection to Q is a $(d-1)$-dimensional surface orthogonal to the velocity vector v whose curvature operator at the point x is $\mathscr{B}^u(x) = \{[\mathscr{B}^u_-(x_+)]^{-1} - \tau(x) I\}^{-1}$, where $\tau(x)$ is the first positive time of reflection of the orbit starting at x and $x_+ = \Phi^{\tau(x)+0}(x) \in M$.

The singularities of the flow $\{\Phi^t\}$ are smooth hypersurfaces in $\mathfrak{M}$. For each finite t there are a finite number of compact smooth hypersurfaces of singularities for Φ^t. The set of singularities for Φ^t for all $t > 0$ (all $t < 0$) consists of a countable number of compact smooth hypersurfaces which we denote S^+ (resp., S^-). It is well known[13,17] that the components of S^+ intersect those of S^- transversally. We claim that, likewise, LUMs $\Gamma^u(x)$, $x \in \mathfrak{M}$, intersect the components of S^+ transversally. We will prove this claim by developing geometry on the singularity sets $S^{\pm}$.

Let $x = (q, v) \in S^+ \cup S^-$ and $n(x)$ be the normal vector to $S^+ \cup S^-$ at x. According to our tradition, we decompose it as $n(x) = (n_q(x), n_v(x))$, with $n_q(x) \in \mathscr{T}_q(Q)$ and $n_v(x) \in \mathscr{T}_v(S^{d-1})$. Since we have identified $\mathscr{T}_v S^{d-1}$ with J_x, the vector $n_v(x)$ belongs in J_x. Moreover, $n_q(x)$ belong in J_x, too, because $S^+ \cup S^-$ is invariant under Φ^t. At a point $x_t = \Phi^t x$, $t \in \mathbb{R}$, we have another normal vector $n(x_t) = (n_q(x_t), n_v(x_t))$ to $S^+ \cup S^-$. The relation between $n(x)$ and $n(x_t)$ can be derived from the following observations:

(i) If no reflection occurs in the interval $(0, t)$, then

$$n_q(x_t) = n_q(x) \qquad \text{and} \qquad n_v(x_t) = n_v(x) - t n_q(x) \tag{A.13}$$

(ii) If t is an instant of reflection at a point $x_t = (q_t, v_t) \in M$, then

$$n_q(x_{t+0}) = n_q(x_{t-0}) - K(x_{t+0})\, n_v(x_{t-0}) \qquad \text{and} \qquad n_v(x_{t+0}) = n_v(x_{t-0}) \tag{A.14}$$

where we use the same notations as in (2.2) and again identify the spaces $J_{x_{t-0}}$ and $J_{x_{t+0}}$. One should note that the relations (A.13)–(A.14) just give a normal vector, not necessarily a unit one. We do not care about the norm of $n(x_t)$, since any restriction on it would complicate our calculations and would not help in any way.

The next quantity we consider is the scalar product $s(x) = (n_q(x), n_v(x))$, which is a function on $S^+ \cup S^-$. Its sign does not depend on the choice of the normal vector. The relations (A.13)–(A.14) readily imply the following

Lemma A.3. If $s(x_t)$ is negative (nonpositive) at $t = t'$, then it remains negative (nonpositive) for all $t \geqslant t'$. Likewise, if it is positive (nonnegative) at $t = t'$, then it is also positive (nonnegative) for all $t \leqslant t'$.

Any point $x = (q, v) \in S_0$ [i.e., any point with $q \in \partial Q$ and $(v, n(q)) = 0$] is, in a sense, an "origin" of both S^+ and S^-. Precisely, $S^+ = \bigcup_{t>0} \Phi^{-t} S_0$ and $S^- = \bigcup_{t>0} \Phi^t S_0$. (Note that since the velocity vectors are tangent to ∂Q on S_0, there is no problem in moving S_0 both forward and backward in time.) It is easily seen that the normal vector $n(x)$ to both S^+ and S^- at any point $x \in S_0$ is $n(x) = (n(q), 0)$. Therefore, $s(x)$ is strictly positive on $S^+ \backslash (S_0)$ and strictly negative on $S^- \backslash (S_0)$. In particular, we get another proof that S^+ and S^- always intersect transversally. A bit more detailed analysis of (A.13)–(A.14) leads to one more conclusion:

Lemma A.4. Apart from a vicinity of the subset S_0 in $\mathfrak{M}$, the vectors $n_q(x)$ and $n_v(x)$ are comparable in length: $c_1 \leqslant \|n_q(x)\| / \|n_v(x)\| \leqslant c_2$, with some positive constants c_1 and c_2 depending only on the vicinity of S_0 that one excludes.

The next lemma follows from the previous one and a known fact that the eigenvalues of $\mathscr{B}^u(x)$, $x \in \mathfrak{M}$, are uniformly bounded away from zero:

Lemma A.5. Apart from a vicinity of $S_0 \subset \mathfrak{M}$, the hypersurfaces of S^+ always intersect LUMs Γ^u of the flow Φ^t at angles $\geqslant c_3$ with a positive constant c_3 depending only on the vicinity of S_0 one excludes.

By the angle between a surface of S^+ and an LUM $\Gamma^u(x)$ intersecting at $x \in S^+$ we naturally mean the angle between the normal vector $n(x)$ to S^+ and the d-dimensional subspace in $\mathcal{T}_x\mathfrak{M}$ orthogonal to $\mathcal{T}_x\Gamma^u(x)$. To prove Lemma A.5, one can easily show that the angle between the vectors $n(x) = (n_q(x), n_v(x))$ and $n_1 = (n_v(x), \mathcal{B}^u(x)\, n_v(x))$ [the latter is tangent to $\Gamma^u(x)$] is uniformly bounded away from $\pi/2$.

Proof of Sublemma 4.1a. First, we notice that for each $m \geqslant 1$ there is a finite number of manifolds in $S_{-m,0}$, and they touch a finite number of compact smooth components of S^+. We denote the union of the latter by S_m^+. The sectional curvature of those components is bounded above by a finite quantity C_m (C_m may depend on m, but it does not matter for us how fast C_m grows with m). To verify this claim, it is enough to show that (i) the sectional curvature of both S^+ and S^- is bounded at the points of their origin, i.e., on S_0, and (ii) the sectional curvature does not grow too rapidly during the evolution of S_0 under Φ^t and Φ^{-t}, $t > 0$, i.e., it stays finite at finite times. Part (i) follows from the above remark that $n(x) = (n(q), 0)$ on S_0; recall that the sectional curvature of ∂Q is bounded. Part (ii) can be derived by a direct calculation based on the "equations of motion" (A.13)–(A.14) for the normal vectors to $\Phi^{\pm t}S_0$ as t grows, and we omit that.

Due to Assumption B, there is an ε_m such that any LUM Γ^u of size $< \varepsilon_m$ intersects no more than K_0 components of S_m^+. Since ε_m depends on m, it can be adjusted to whatever large value of C_m, and then we can think of those components as almost flat hypersurfaces in a vicinity of Γ^u. They intersect the HLUM transversally by virtue of Lemma A.5.

We now consider an HLUM $\gamma_1^u \subset M$ of a ρ size $< \varepsilon_m$. The structure of the ρ metric on γ_1^u can be better understood if one considers an orthogonal cross section Σ^u of a bundle of trajectories coming to the set $\gamma_1^u \subset M$, just before the reflection. The natural Riemannian metric in Σ^u is isomorphic to ρ in γ_1^u. By equipping the surface Σ^u with unit normal vectors pointing in the flow direction, we get an HLUM Γ_1^u for the flow. Since γ_1^u is cut by $\leqslant K_0$ components of $S_{-m,0}$, the HLUM Γ_1^u is cut by $\leqslant K_0$ hypersurfaces of S_m^+.

We denote by ρ_1 the normalized ρ measure on γ_1^u and by $\langle \cdot \rangle_1$ the expectation with respect to ρ_1. To emulate the functions $r_n(x)$ and $r_n'(x)$ from Sublemma 4.1a, we introduce two functions on Σ^u: $r(x) = \mathrm{dist}(x, \partial\Sigma^u)$ and $r'(x) = \mathrm{dist}(x, \partial\Sigma^u \cup (S_m^+ \cap \Sigma^u))$. We claim that

$$-\langle \ln r'(x) \rangle_1 \leqslant -\langle \ln r(x) \rangle_1 + C' \cdot K_0 \qquad \text{(A.15)}$$

provided ε_m is small enough; here C' is independent of m or γ_1^u. Evidently, it suffices to prove (A.15) for $K_0 = 1$. Since the sectional curvature of Σ^u is uniformly bounded and ε_m is small enough, the surface Σ^u is almost flat. One can think of it as just a domain in $\mathbb{R}^{d-1}$ and of ρ as just the Euclidean metric (then ρ_1 is just the normalized Lebesgue measure). Likewise, the cutting surface (recall that $K_0 = 1$) is almost flat, and one can think of it as a hyperplane cutting the above domain. After that the problem boils down to a rather simple geometric consideration. We take an arbitrary segment Z orthogonal to the cutting hyperplane whose endpoints belob to the boundary of the domain. The conditional ρ measure on that segment is proportional the Lebesgue measure (length). We denote by $\langle \cdot \rangle_Z$ the expectation with respect to that measure. The condition $r'(x) < r(x)$ holds on a subsegment that is at least twice as short as Z. Then it is an elementary calculation that $\langle \ln r'(x) \rangle_Z < \langle \ln r(x) \rangle_Z + \ln 2$. Integrating over the domain gives (A.15) with $C' = \ln 2$.

Finally, we recall that the HLUM γ_1^u is supposed to be a component of the image of the original HLUM γ^u in Lemma 4.1. Hence the normalized measure ρ_2 on γ_1^u induced by pulling ρ_0 from γ^u differs from the existing ρ_1 measure on γ_1^u. However, the Radon–Nikodým derivative $d\rho_2(x)/d\rho_1(x)$ is uniformly bounded away from zero and infinity by virtue of (3.8). Thus, we get $-\langle \ln r'(x) \rangle_2 \leqslant -\langle \ln r(x) \rangle_2 + C'' \cdot K_0$, where the expectation is taken with respect to ρ_2, with some C'' determined by C' and C_0 in (3.8). Sublemma 4.1a is proven. ■

Remark A.6. The last step in the proof of (A.15) was the integration over the domain representing γ_1^u. We now suppose that the functions $r(x)$ and $r'(x)$ only differ on a relatively small subdomain. Then one can strengthen (A.15) as

$$-\langle \ln r'(x) \rangle_1 \leqslant -\langle \ln r(x) \rangle_1 + PC'K_0 \tag{A.16}$$

where $P = 2\rho_1(\{x \in \gamma_1^u : r(x) \neq r'(x)\})$. We now can support the bound (4.1). The values $r_n^{\text{new}}(x)$ and $r_n^{\text{old}}(x)$ can differ only in the $2D$ neighborhood of the boundary $\partial\gamma_{1,s}^u$. This neighborhood obviously has the relative ρ measure $\leqslant \text{const} \cdot P^s/P^v$ in $\gamma_{1,v}^u$ in the notations used in (4.1). Thus, (4.1) follows from (A.16).

Remark A.7. In a similar fashion one can obtain (4.3). Here the key observation is that $r_n(x) < 2D$ on any moving subcomponent (no matter how long it is). Thus, the additional cuttings defined in Section 4 can only alter the function $r_n(x)$ in the $2D$ neighborhood of the cutting surfaces. There is a certain freedom in positioning those surfaces and one can minimize the ρ measure of their $2D$ neighborhood, so that its relative ρ measure in the whole moving component will be less than, say, $100D/\varepsilon_m$. The bound (4.3) then follows from (A.16).

Proof of Sublemma 4.1B. Consider an HLUM γ_1^u which intersects some hypersurfaces of $\mathscr{D}_0$. Again, as in the proof of Sublemma 4.1a, we think of γ_1^u as a "flat" $(d-1)$-dimensional domain cut by some hyperplanes. Let $\mathscr{H}_k$ and $\mathscr{H}_{k+1}$ be two neighboring hyperplanes defined by the equations $(v, n(q)) = k^{-\theta}$ and $(v, n(q)) = (k+1)^{-\theta}$, respectively. Denote the part of γ_1^u confined between $\mathscr{H}_k$ and $\mathscr{H}_{k+1}$ by $\gamma_{1,k}^u$. The ρ distance between $\mathscr{H}_k$ and $\mathscr{H}_{k+1}$ in $\gamma_{1,k}^u$ is easily seen to be of order $k^{-2\theta-1}$ (this means that it is between $c_1 k^{-2\theta-1}$ and $c_2 k^{-2\theta-1}$ with some constants c_1 and c_2 independent of k or γ_1^u).

We consider an arbitrary segment Z in $\gamma_{1,k}^u$ whose endpoints belong to hyperplanes $\mathscr{H}_k$ and $\mathscr{H}_{k+1}$, and which is perpendicular to one of them, say, to $\mathscr{H}_k$. For each $x \in Z$ denote $r(x)$ the ρ distance of x from $\partial\gamma_1^u$. Note that γ_1^u does not intersect the singularity set S_0 and the ρ distance from any $x \in Z$ to S_0 is of order $k^{-2\theta}$. Hence, $r(x) \leqslant \text{const} \cdot k^{-2\theta}$.

The map T is almost linear on $\gamma_{1,k}^u$, and the image $T\gamma_{1,k}^u$ is also an almost flat $(d-1)$-dimensional compact surface in M (provided ε_m is small enough). The distance between $T\mathscr{H}_k$ and $T\mathscr{H}_{k+1}$ in the ρ metric in $T\gamma_{1,k}^u$ is easily seen to be of order $k^{-\theta-1}$, because the rate of expansion under T is proportional to $(v, n(q))^{-1} \approx k^{\theta}$. For every point $x \in Z$ we denote by $r'(x)$ the ρ distance of Tx from the boundary $\partial(T\gamma_{1,k}^u)$ and by $r''(x)$ the ρ distance of Tx from $T(\mathscr{H}_k \cup \mathscr{H}_{k+1})$. Let $Z_1 = \{x \in Z\colon r'(x) < r''(x)\}$ and $Z_2 = Z \backslash Z_1$. Denote by $l(\cdot)$ the normalized Lebesgue measure (length) on Z and by $\langle \cdot \rangle_Z$ the expectation with respect to $l(\cdot)$. For each $x \in Z_1$ we have $r'(x) \geqslant \Lambda_0 r(x)$ with a constant $\Lambda_0 > 1$ due to uniform expansion on LUMs. Thus, one has

$$-\int_{Z_1} \ln r'(x)\, dl(x) \leqslant -\int_{Z_1} \ln r(x)\, dl(x) - l(Z_1) \ln \Lambda_0 \tag{A.17}$$

On the other hand, one has $-\ln r(x) \geqslant \text{const} + 2\theta \ln k$ for each $x \in Z$ and $l\{x \in Z_2\colon r''(x) < \delta\} \leqslant \text{const} \cdot \delta k^{\theta+1}$ for any $\delta > 0$. As a result, one obtains

$$\begin{aligned} &-\int_{Z_2} \ln r''(x)\, dl(x) \\ &\quad\leqslant -\int_{Z^2} \ln r(x)\, dl(x) - l(Z_2)[(\theta - 1)\ln k - \text{const}] - l(Z_2) \ln l(Z_2) \end{aligned} \tag{A.18}$$

Since $k \geqslant n_0$ and n_0 is supposed to be large enough, adding (A.17) and (A.18) gives a bound

$$-\int_Z \ln r'(x)\, dl(x) \leqslant -\int_Z \ln r(x)\, dl(x) - \tfrac{1}{2} \ln \Lambda_0$$

The proof of Sublemma 4.1b is then accomplished by integration over γ_1^u and by using (3.8) as in the proof of Sublemma 4.1a.

ACKNOWLEDGMENTS

I am deeply indebted to Ya. G. Sinai for stimulating my interest in the subject. I thank L. A. Bunimovich for enlightening discussions. This work was done during my visit at the University of California (Los Angeles), and I should like to thank L.-S. Young and P. Fan for their kind hospitality and K. Hirvasaho for her invaluable help in editing my English. I also thank the referees for their criticism on the first version of the manuscript, which encouraged me to include more detailed proofs and add an appendix.

REFERENCES

1. P. M. Bleher, Statistical properties of two-dimensional periodic Lorentz gas with infinite horizon, *J. Stat. Phys.* **66**:315–373 (1992).
2. J.-P. Bouchaud and P. Le Doussal, Numerical study of a *d*-dimensional periodic Lorentz gas with universal properties, *J. Stat. Phys.* **41**:225–248 (1985).
3. R. Bowen, *Equilibrium States and Ergodic Theory of Anosov Diffeomorphisms*, (Springer, Berlin, 1975).
4. R. Bowen, Markov partitions are not smooth, *Proc. Am. Math. Soc.* **71**:130–132 (1978).
5. L. A. Bunimovich and Ya. G. Sinai, Markov partitions for dispersed billiards, *Commun. Math. Phys.* **73**:247–280 (1980).
6. L. A. Bunimovich and Ya. G. Sinai, Statistical properties of Lorentz gas with periodic configuration of scatterers, *Commun. Math. Phys.* **78**:479–497 (1981).
7. L. A. Bunimovich, Ya. G. Sinai, and N. I. Chernov, Statistical properties of two-dimensional hyperbolic billiards, *Russ. Math. Surv.* **46**:47–106 (1991).
8. N. I. Chernov, Ergodic and statistical properties of piecewise linear hyperbolic automorphisms of the 2-torus, *J. Stat. Phys.* **69**:111–134 (1992).
9. N. I. Chernov, G. L. Eyink, J. L. Lebowitz, and Ya. G. Sinai, Steady-state electrical conduction in the periodic Lorentz gas, *Commun. Math. Phys.* **154**:569–601 (1993).
10. G. Gallavotti and D. Ornstein, Billiards and Bernoulli schemes, *Commun. Math. Phys.* **38**:83–101 (1974).
11. I. A. Ibragimov and Y. V. Linnik, *Independent and Stationary Sequences of Random Variables* (Wolters-Noordhoff, Gröningen, 1971).
12. A. Katok and J.-M. Strelcyn, *Invariant Manifolds, Entropy and Billiards; Smooth Maps with Singularities* (Springer, New York, 1986).
13. A. Krámli, N. Simányi, and D. Szász, A "transversal" fundamental theorem for semi-dispersing billiards, *Commun. Math. Phys.* **129**:535–560 (1990).
14. T. Krüger and S. Troubetzkoy, Markov partitions and shadowing for nonuniformly hyperbolic systems with singularities, *Ergodic Theory Dynamic Syst.* **12**:487–508 (1992).
15. Ya. G. Sinai, Construction of Markov partitions, *Funkts. Anal. Ego Prilozhen.* **2**:70–80 (1968).
16. Ya. G. Sinai, Dynamical systems with elastic reflections. Ergodic properties of dispersing billiards, *Russ. Math. Surv.* **25**:137–189 (1970).
17. Ya. G. Sinai and N. I. Chernov, Ergodic properties of some systems of 2-dimensional discs and 3-dimensional spheres, *Russ. Math. Surv.* **42**:181–207 (1987).