July 22, 2004

The Boltzmann-Sinai Ergodic Hypothesis in Two Dimensions (Without Exceptional Models)

Nándor Simányi[1]

University of Alabama at Birmingham
Department of Mathematics
Campbell Hall, Birmingham, AL 35294 U.S.A.
E-mail: simanyi@math.uab.edu

Abstract. We consider the system of N (≥ 2) elastically colliding hard balls of masses $m_1, \dots, m_N$ and radius r in the flat unit torus $\mathbb{T}^\nu$, $\nu \geq 2$. In the case $\nu = 2$ we prove (the full hyperbolicity and) the ergodicity of such systems for every selection $(m_1, \dots, m_N; r)$ of the external geometric parameters, without exceptional values. In higher dimensions, for hard ball systems in $\mathbb{T}^\nu$ ($\nu \geq 3$), we prove that every such system (is fully hyperbolic and) has open ergodic components.

Primary subject classification: 37D50

Secondary subject classification: 34D05

§1. Introduction

This paper contains a step toward a complete solution (yet to be achieved) of the celebrated Boltzmann–Sinai ergodic hypothesis. In a loose form, as attributed to L. Boltzmann back in the 1880's, it asserts that gases of hard balls are ergodic. In a precise form, which is due to Ya. Sinai in 1963 [Sin(1963)], it states that the gas of $N \geq 2$ identical hard balls (of small radius) on a torus $\mathbb{T}^\nu$, $\nu \geq 2$, (a ν-dimensional box with periodic boundary conditions) is ergodic, provided that certain necessary

[1]Research supported by the National Science Foundation, grant DMS-0098773.

reductions have been made. The latter means that one fixes the total energy, sets the total momentum to zero, and restricts the center of mass to a certain discrete lattice within the torus. The assumption of a small radius is necessary to have the configuration space connected.

Sinai himself pioneered rigorous mathematical studies of hard ball gases by proving the hyperbolicity and ergodicity for the case $N = 2$ and $\nu = 2$ in his seminal paper [Sin(1970)], where he laid down the foundation of the modern theory of chaotic billiards. Then Chernov and Sinai extended this result to ($N = 2$, $\nu > 2$), as well as proved a general theorem on "local" ergodicity applicable to systems of $N > 2$ balls [S-Ch(1987)]; the latter became instrumental in the subsequent studies. The case $N > 2$ is substantially more difficult than that of $N = 2$ because, while the system of two balls reduces to a billiard with strictly convex (spherical) boundary, which guarantees strong hyperbolicity, the gases of $N > 2$ balls reduce to billiards with convex, but not strictly convex, boundary (the latter is a finite union of cylinders) – those are characterized by very weak hyperbolicity.

Further development has been mostly due to A. Krámli, D. Szász, and the present author. We proved hyperbolicity and ergodicity for $N = 3$ balls in any dimension [K-S-Sz(1991)] by employing the "local" ergodic theorem of Chernov and Sinai, and carefully analyzing all possible degeneracies in the dynamics to obtain "global" ergodicity. We extended our results to $N = 4$ balls in dimension $\nu \geq 3$ next year, and then I proved the ergodicity whenever $N \leq \nu$ (this covers systems with an arbitrary number of balls, but only in spaces of high enough dimension, which is a restrictive condition). At this point, the existing methods could no longer handle any new cases, because the analysis of the degeneracies became overly complicated. It was clear that further progress should involve novel ideas.

A breakthrough was achieved by Szász and myself, when we employed the methods of algebraic geometry [S-Sz(1999)]. We assumed that the balls had arbitrary masses $m_1, \dots, m_N$ (but the same radius r). Now by taking the limit $m_N \to 0$, we were able to reduce the dynamics of N balls to the motion of $N - 1$ balls, thus utilizing a natural induction on N. Then algebro-geometric methods allowed us to effectively analyze all possible degeneracies, but only for typical (generic) vectors of "external" parameters $(m_1, \dots, m_N, r)$; the latter needed to avoid some exceptional submanifolds of codimension one, which remained unknown. This approach led to a proof of full hyperbolicity (but not yet ergodicity) for all $N \geq 2$ and $\nu \geq 2$, and for generic $(m_1, \dots, m_N, r)$, see [S-Sz(1999)]. Later I significantly simplified the arguments and made them more "dynamical", which allowed me to obtain full hyperbolicity for hard balls with any set of external geometric parameters $(m_1, \dots, m_N, r)$ [Sim(2002)]. Thus, the hyperbolicity has been fully established for all systems of hard balls on a torus.

To upgrade the full hyperbolicity to ergodicity one needs to refine the analysis of the aforementioned degeneracies. For hyperbolicity, it was enough that the degeneracies made a subset of codimension ≥ 1 in the phase space. For ergodicity, one has to show that its codimension is ≥ 2. In the paper [Sim(2003)] I took the first step in this direction – I proved that systems of $N \geq 2$ balls on a 2D torus (i.e., $\nu = 2$) are ergodic for typical (generic) vectors of external parameters $(m_1, \dots, m_N, r)$. The proof again involves some algebro-geometric techniques, thus the result is restricted to "generic" masses. But there is a good reason to believe that systems in $\nu \geq 3$ dimensions would be somewhat easier to handle, at least that was indeed the case in early studies.

In this paper I will prove two theorems:

Theorem 1. In the case $\nu = 2$, for any integer value $N \geq 2$ and for every $(N+1)$-tuple $(m_1, \dots, m_N, r)$ of the external geometric parameters the standard hard ball system $\left(\mathbf{M}_{\vec{m},r}, \left\{S^t_{\vec{m},r}\right\}, \mu_{\vec{m},r}\right)$ is (fully hyperbolic and) ergodic.

Theorem 2. In the case $\nu \geq 3$, for any integer value $N \geq 2$ and for every $(N+1)$-tuple $(m_1, \dots, m_N, r)$ of the external geometric parameters the standard hard ball system $\left(\mathbf{M}_{\vec{m},r}, \left\{S^t_{\vec{m},r}\right\}, \mu_{\vec{m},r}\right)$ (is fully hyperbolic and) has *open* ergodic components.

Remark 1.1. The novelty of Theorem 1 (as compared to the results in [Sim(2003)]) is that it applies to each $(N+1)$-tuple of external parameters (provided that the interior of the phase space is connected), without an exceptional zero-measured set.

Remark 1.2. Theorem 2 speaks about exactly the same models as the result of [Sim(2002)], but the assertion of this new theorem is stronger than that of the theorem in [Sim(2002)]: Theorem 2 guarantees open ergodic components, not only positively measured ones. (The latter is a standard corollary of the full hyperbolicity in semi-dispersive billiards.)

Remark 1.3. As it follows from the results of [C-H(1996)] and [O-W(1998)], every two-dimensional ($\nu = 2$), standard hard ball system $(\mathbf{M}, \{S^t\}_{t\in\mathbb{R}}, \mu)$ (the models covered by Theorem 1) is not only ergodic, but enjoys the Bernoulli mixing property, as well. Likewise, in the case $\nu \geq 3$, every ν-dimensional, standard hard ball flow appears to be B-mixing on each of its open ergodic components.

§2. Prerequisites

Consider the ν-dimensional ($\nu \geq 2$), standard, flat torus $\mathbb{T}^\nu = \mathbb{R}^\nu/\mathbb{Z}^\nu$ as the vessel containing N (≥ 2) hard balls (spheres) $B_1, \dots, B_N$ with positive masses $m_1, \dots, m_N$ and (just for simplicity) common radius $r > 0$. We always assume that the radius $r > 0$ is not too big, so that even the interior of the arising configuration space $\mathbf{Q}$ (or, equivalently, the phase space) is connected. Denote the center of the ball B_i by $q_i \in \mathbb{T}^\nu$, and let $v_i = \dot{q}_i$ be the velocity of the i-th particle. We investigate the uniform motion of the balls $B_1, \dots, B_N$ inside the container $\mathbb{T}^\nu$ with half a unit of total kinetic energy: $E = \frac{1}{2}\sum_{i=1}^N m_i ||v_i||^2 = \frac{1}{2}$. We assume that the collisions between balls are perfectly elastic. Since — beside the kinetic energy E — the total momentum $I = \sum_{i=1}^N m_i v_i \in \mathbb{R}^\nu$ is also a trivial first integral of the motion, we make the standard reduction $I = 0$. Due to the apparent translation invariance of the arising dynamical system, we factorize the configuration space with respect to uniform spatial translations as follows: $(q_1, \dots, q_N) \sim (q_1 + a, \dots, q_N + a)$ for all translation vectors $a \in \mathbb{T}^\nu$. The configuration space $\mathbf{Q}$ of the arising flow is then the factor torus $\left((\mathbb{T}^\nu)^N / \sim \right) \cong \mathbb{T}^{\nu(N-1)}$ minus the cylinders

$$C_{i,j} = \left\{ (q_1, \dots, q_N) \in \mathbb{T}^{\nu(N-1)} \colon \operatorname{dist}(q_i, q_j) < 2r \right\}$$

$(1 \le i < j \le N)$ corresponding to the forbidden overlap between the i-th and j-th spheres. Then it is easy to see that the compound configuration point

$$q = (q_1, \dots, q_N) \in \mathbf{Q} = \mathbb{T}^{\nu(N-1)} \setminus \bigcup_{1 \le i < j \le N} C_{i,j}$$

moves in $\mathbf{Q}$ uniformly with unit speed and bounces back from the boundaries $\partial C_{i,j}$ of the cylinders $C_{i,j}$ according to the classical law of geometric optics: the angle of reflection equals the angle of incidence. More precisely: the post-collision velocity v^+ can be obtained from the pre-collision velocity v^- by the orthogonal reflection across the tangent hyperplane of the boundary $\partial \mathbf{Q}$ at the point of collision. Here we must emphasize that the phrase "orthogonal" should be understood with respect to the natural Riemannian metric (the kinetic energy) $||dq||^2 = \sum_{i=1}^N m_i ||dq_i||^2$ in the configuration space $\mathbf{Q}$. For the normalized Liouville measure μ of the arising flow $\{S^t\}$ we obviously have $d\mu = \text{const} \cdot dq \cdot dv$, where dq is the Riemannian volume in $\mathbf{Q}$ induced by the above metric, and dv is the surface measure (determined by the restriction of the Riemannian metric above) on the unit sphere of compound velocities

$$\mathbb{S}^{\nu(N-1)-1} = \left\{ (v_1, \dots, v_N) \in (\mathbb{R}^\nu)^N \colon \sum_{i=1}^N m_i v_i = 0 \text{ and } \sum_{i=1}^N m_i ||v_i||^2 = 1 \right\}.$$

The phase space $\mathbf{M}$ of the flow $\{S^t\}$ is the unit tangent bundle $\mathbf{Q} \times \mathbb{S}^{d-1}$ of the configuration space $\mathbf{Q}$. (We will always use the shorthand notation $d = \nu(N-1)$ for the dimension of the billiard table $\mathbf{Q}$.) We must, however, note here that at the boundary $\partial \mathbf{Q}$ of $\mathbf{Q}$ one has to glue together the pre-collision and post-collision velocities in order to form the phase space $\mathbf{M}$, so $\mathbf{M}$ is equal to the unit tangent bundle $\mathbf{Q} \times \mathbb{S}^{d-1}$ modulo this identification.

A bit more detailed definition of hard ball systems with arbitrary masses, as well as their role in the family of cylindric billiards, can be found in §4 of [S-Sz(2000)] and in §1 of [S-Sz(1999)]. We denote the arising flow by $(\mathbf{M}, \{S^t\}_{t\in\mathbb{R}}, \mu)$.

In the series of articles [K-S-Sz(1989)], [K-S-Sz(1991)], [K-S-Sz(1992)], [Sim(1992-I)], and [Sim(1992-II)] the authors developed a powerful, three-step strategy for proving the (hyperbolic) ergodicity of hard ball systems. First of all, these proofs are inductions on the number N of balls involved in the problem. Secondly, the induction step itself consists of the following three major steps:

Step I. To prove that every non-singular (i. e. smooth) trajectory segment $S^{[a,b]}x_0$ with a "combinatorially rich" (in a well defined sense) symbolic collision sequence is automatically sufficient (or, in other words, "geometrically hyperbolic", see below in this section), provided that the phase point x_0 does not belong to a countable union J of smooth sub-manifolds with codimension at least two. (Containing the exceptional phase points.)

The exceptional set J featuring this result is negligible in our dynamical considerations — it is a so called slim set. For the basic properties of slim sets, see again below in this section.

Step II. Assume the induction hypothesis, i. e. that all hard ball systems with N' balls ($2 \le N' < N$) are (hyperbolic and) ergodic. Prove that there exists a slim set $S \subset \mathbf{M}$ with the following property: For every phase point $x_0 \in \mathbf{M} \setminus S$ the entire trajectory $S^{\mathbb{R}}x_0$ contains at most one singularity and its symbolic collision sequence is combinatorially rich, just as required by the result of Step I.

Step III. By using again the induction hypothesis, prove that almost every singular trajectory is sufficient in the time interval $(t_0, +\infty)$, where t_0 is the time moment of the singular reflection. (Here the phrase "almost every" refers to the volume defined by the induced Riemannian metric on the singularity manifolds.)

We note here that the almost sure sufficiency of the singular trajectories (featuring Step III) is an essential condition for the proof of the celebrated Theorem on Local Ergodicity for algebraic semi-dispersive billiards proved by Bálint–Chernov–Szász–Tóth in [B-Ch-Sz-T (2002)]. Under this assumption Theorem 4.4 of [B-Ch-Sz-T (2002)] states that in any algebraic semi-dispersive billiard system (i. e. in a system such that the smooth components of the boundary $\partial\mathbf{Q}$ are algebraic hypersurfaces) a suitable, open neighborhood U_0 of any sufficient phase point $x_0 \in \mathbf{M}$ (with at most one singularity on its trajectory) belongs to a single ergodic component of the billiard flow $(\mathbf{M}, \{S^t\}_{t\in\mathbb{R}}, \mu)$.

In an inductive proof of ergodicity, steps I and II together ensure that there exists an arc-wise connected set $C \subset \mathbf{M}$ with full measure, such that every phase point $x_0 \in C$ is sufficient with at most one singularity on its trajectory. Then the cited Theorem on Local Ergodicity (now taking advantage of the result of Step III) states that for every phase point $x_0 \in C$ an open neighborhood U_0 of x_0 belongs to one ergodic component of the flow. Finally, the connectedness of the set C and

$\mu(\mathbf{M}\setminus C)=0$ imply that the flow $(\mathbf{M},\{S^t\}_{t\in\mathbb{R}},\mu)$ (now with N balls) is indeed ergodic, and actually fully hyperbolic, as well.

The generator subspace $A_{i,j}\subset\mathbb{R}^{\nu N}$ $(1\le i<j\le N)$ of the cylinder $C_{i,j}$ (describing the collisions between the i-th and j-th balls) is given by the equation

$$A_{i,j}=\left\{(q_1,\dots,q_N)\in(\mathbb{R}^\nu)^N:\ q_i=q_j\right\}, \tag{2.1}$$

see (4.3) in [S-Sz(2000)]. Its ortho-complement $L_{i,j}\subset\mathbb{R}^{\nu N}$ is then defined by the equation

$$L_{i,j}=\left\{(q_1,\dots,q_N)\in(\mathbb{R}^\nu)^N:\ q_k=0 \text{ for } k\ne i,j, \text{ and } m_iq_i+m_jq_j=0\right\}, \tag{2.2}$$

see (4.4) in [S-Sz(2000)]. Easy calculation shows that the cylinder $C_{i,j}$ (describing the overlap of the i-th and j-th balls) is indeed spherical and the radius of its base sphere is equal to $r_{i,j}=2r\sqrt{\frac{m_im_j}{m_i+m_j}}$, see §4, especially formula (4.6) in [S-Sz(2000)].

The structure lattice $\mathcal{L}\subset\mathbb{R}^{\nu N}$ is clearly the lattice $\mathcal{L}=(\mathbb{Z}^\nu)^N=\mathbb{Z}^{N\nu}$.

Due to the presence of an additional invariant quantity $I=\sum_{i=1}^N m_iv_i$, one usually makes the reduction $\sum_{i=1}^N m_iv_i=0$ and, correspondingly, factorizes the configuration space with respect to uniform spatial translations:

$$(q_1,\dots,q_N)\sim(q_1+a,\dots,q_N+a),\quad a\in\mathbb{T}^\nu. \tag{2.3}$$

The natural, common tangent space of this reduced configuration space is

$$\mathcal{Z}=\left\{(v_1,\dots,v_N)\in(\mathbb{R}^\nu)^N:\ \sum_{i=1}^N m_iv_i=0\right\}=\left(\bigcap_{i<j}A_{i,j}\right)^\perp=(\mathcal{A})^\perp \tag{2.4}$$

supplied with the inner product

$$\langle v,\,v'\rangle=\sum_{i=1}^N m_i\langle v_i,\,v_i'\rangle,$$

see also (4.1) and (4.2) in [S-Sz(2000)].

Collision graphs. Let $S^{[a,b]}x$ be a nonsingular, finite trajectory segment with the collisions $\sigma_1,\dots,\sigma_n$ listed in time order. (Each σ_k is an unordered pair (i,j) of different labels $i,j\in\{1,2,\dots,N\}$.) The graph $\mathcal{G}=(\mathcal{V},\mathcal{E})$ with vertex set $\mathcal{V}=\{1,2,\dots,N\}$ and set of edges $\mathcal{E}=\{\sigma_1,\dots,\sigma_n\}$ is called the *collision graph* of the orbit segment $S^{[a,b]}x$. For a given positive number C, the collision graph $\mathcal{G}=(\mathcal{V},\mathcal{E})$ of the orbit segment $S^{[a,b]}x$ will be called *C-rich* if $\mathcal{G}$ contains at least C connected, consecutive (i. e. following one after the other in time, according to the time-ordering given by the trajectory segment $S^{[a,b]}x$) subgraphs.

Trajectory Branches. We are going to briefly describe the discontinuity of the flow $\{S^t\}$ caused by a multiple collisions at time t_0. Assume first that the pre–collision velocities of the particles are given. What can we say about the possible post–collision velocities? Let us perturb the pre–collision phase point (at time $t_0 - 0$) infinitesimally, so that the collisions at $\sim t_0$ occur at infinitesimally different moments. By applying the collision laws to the arising finite sequence of collisions, we see that the post-collision velocities are fully determined by the time-ordered list of the arising collisions. Therefore, the collection of all possible time-ordered lists of these collisions gives rise to a finite family of continuations of the trajectory beyond t_0. They are called the trajectory branches. It is quite clear that similar statements can be said regarding the evolution of a trajectory through a multiple collision in reverse time. Furthermore, it is also obvious that for any given phase point $x_0 \in \mathbf{M}$ there are two, ω-high trees $\mathcal{T}_+$ and $\mathcal{T}_-$ such that $\mathcal{T}_+$ ($\mathcal{T}_-$) describes all the possible continuations of the positive (negative) trajectory $S^{[0,\infty)}x_0$ ($S^{(-\infty,0]}x_0$). (For the definitions of trees and for some of their applications to billiards, cf. the beginning of §5 in [K-S-Sz(1992)].) It is also clear that all possible continuations (branches) of the whole trajectory $S^{(-\infty,\infty)}x_0$ can be uniquely described by all pairs (B_-, B_+) of infinite branches of the trees $\mathcal{T}_-$ and $\mathcal{T}_+$ ($B_- \subset \mathcal{T}_-, B_+ \subset \mathcal{T}_+$).

Finally, we note that the trajectory of the phase point x_0 has exactly two branches, provided that $S^t x_0$ hits a singularity for a single value $t = t_0$, and the phase point $S^{t_0}x_0$ does not lie on the intersection of more than one singularity manifolds. In this case we say that the trajectory of x_0 has a "simple singularity".

Neutral Subspaces, Advance, and Sufficiency. Consider a nonsingular trajectory segment $S^{[a,b]}x$. Suppose that a and b are not moments of collision.

Definition 2.5. *The neutral space $\mathcal{N}_0(S^{[a,b]}x)$ of the trajectory segment $S^{[a,b]}x$ at time zero ($a < 0 < b$) is defined by the following formula:*

$$\begin{aligned}&\mathcal{N}_0(S^{[a,b]}x) = \big\{W \in \mathcal{Z}\colon \exists(\delta > 0) \ \textit{ such that } \ \forall\alpha \in (-\delta, \delta)\\ &V\left(S^a\left(Q(x) + \alpha W, V(x)\right)\right) = V(S^a x) \ \textit{and} \ V\left(S^b\left(Q(x) + \alpha W, V(x)\right)\right) = V(S^b x)\big\}.\end{aligned}$$

($\mathcal{Z}$ is the common tangent space $\mathcal{T}_q\mathbf{Q}$ of the parallelizable manifold $\mathbf{Q}$ at any of its points q, while $V(x)$ is the velocity component of the phase point $x = (Q(x),\, V(x))$.)

It is known (see (3) in §3 of [S-Ch (1987)]) that $\mathcal{N}_0(S^{[a,b]}x)$ is a linear subspace of $\mathcal{Z}$ indeed, and $V(x) \in \mathcal{N}_0(S^{[a,b]}x)$. The neutral space $\mathcal{N}_t(S^{[a,b]}x)$ of the segment $S^{[a,b]}x$ at time $t \in [a,b]$ is defined as follows:

$$\mathcal{N}_t(S^{[a,b]}x) = \mathcal{N}_0\left(S^{[a-t,b-t]}(S^t x)\right).$$

It is clear that the neutral space $\mathcal{N}_t(S^{[a,b]}x)$ can be canonically identified with $\mathcal{N}_0(S^{[a,b]}x)$ by the usual identification of the tangent spaces of $\mathbf{Q}$ along the trajectory $S^{(-\infty,\infty)}x$ (see, for instance, §2 of [K-S-Sz(1990)]).

Our next definition is that of the advance. Consider a non-singular orbit segment $S^{[a,b]}x$ with the symbolic collision sequence $\Sigma = (\sigma_1, \dots, \sigma_n)$, meaning that $S^{[a,b]}x$ has exactly n collisions with $\partial\mathbf{Q}$, and the i-th collision ($1 \le i \le n$) takes place at the boundary of the cylinder C_{σ_i}. For $x = (Q, V) \in \mathbf{M}$ and $W \in \mathcal{Z}$, $\|W\|$ sufficiently small, denote $T_W(Q, V) := (Q + W, V)$.

Definition 2.6. *For any $1 \le k \le n$ and $t \in [a, b]$, the advance*

$$\alpha_k = \alpha(\sigma_k)\colon\ \mathcal{N}_t(S^{[a,b]}x) \to \mathbb{R}$$

of the collision σ_k is the unique linear extension of the linear functional $\alpha_k = \alpha(\sigma_k)$ defined in a sufficiently small neighborhood of the origin of $\mathcal{N}_t(S^{[a,b]}x)$ in the following way:

$$\alpha(\sigma_k)(W) := t_k(x) - t_k(S^{-t}T_W S^t x).$$

Here $t_k = t_k(x)$ is the time moment of the k-th collision σ_k on the trajectory of x after time $t = a$. The above formula and the notion of the advance functional

$$\alpha_k = \alpha(\sigma_k) :\ \mathcal{N}_t\left(S^{[a,b]}x\right) \to \mathbb{R}$$

has two important features:

(i) If the spatial translation $(Q, V) \mapsto (Q + W, V)$ $(W \in \mathcal{N}_t\left(S^{[a,b]}x\right))$ is carried out at time t, then t_k changes linearly in W, and it takes place just $\alpha_k(W)$ units of time earlier. (This is why it is called "advance".)

(ii) If the considered reference time t is somewhere between t_{k-1} and t_k, then the neutrality of W with respect to σ_k precisely means that

$$W - \alpha_k(W) \cdot V(x) \in A_{\sigma_k},$$

i. e. a neutral (with respect to the collision σ_k) spatial translation W with the advance $\alpha_k(W) = 0$ means that the vector W belongs to the generator space A_{σ_k} of the cylinder C_{σ_k}.

It is now time to bring up the basic notion of sufficiency (or, sometimes it is also called geometric hyperbolicity) of a trajectory (segment). This is the utmost important necessary condition for the proof of the fundamental theorem for algebraic semi-dispersive billiards, i. e. Theorem 4.4 in [B-Ch-Sz-T(2002)].

Definition 2.7.

(1) *The nonsingular trajectory segment $S^{[a,b]}x$ (a and b are supposed not to be moments of collision) is said to be sufficient if and only if the dimension of $\mathcal{N}_t(S^{[a,b]}x)$ ($t \in [a, b]$) is minimal, i.e. dim $\mathcal{N}_t(S^{[a,b]}x) = 1$.*

(2) *The trajectory segment $S^{[a,b]}x$ containing exactly one singularity (a so called "simple singularity", see above) is said to be sufficient if and only if both branches of this trajectory segment are sufficient.*

Definition 2.8. *The phase point $x \in \mathbf{M}$ with at most one (simple) singularity is said to be sufficient if and only if its whole trajectory $S^{(-\infty,\infty)}x$ is sufficient, which means, by definition, that some of its bounded segments $S^{[a,b]}x$ are sufficient.*

In the case of an orbit $S^{(-\infty,\infty)}x$ with a simple singularity, sufficiency means that both branches of $S^{(-\infty,\infty)}x$ are sufficient.

No accumulation (of collisions) in finite time. By the results of Vaserstein [V(1979)], Galperin [G(1981)] and Burago-Ferleger-Kononenko [B-F-K(1998)], in any semi-dispersive billiard flow there can only be finitely many collisions in finite time intervals, see Theorem 1 in [B-F-K(1998)]. Thus, the dynamics is well defined as long as the trajectory does not hit more than one boundary components at the same time.

Slim sets. We are going to summarize the basic properties of codimension-two subsets A of a connected, smooth manifold M with a possible boundary and corners. Since these subsets A are just those negligible in our dynamical discussions, we shall call them slim. As to a broader exposition of the issues, see [E(1978)] or §2 of [K-S-Sz(1991)].

Note that the dimension $\dim A$ of a separable metric space A is one of the three classical notions of topological dimension: the covering (Čech-Lebesgue), the small inductive (Menger-Urysohn), or the large inductive (Brouwer-Čech) dimension. As it is known from general topology, all of them are the same for separable metric spaces.

Definition 2.9. A subset A of M is called slim if and only if A can be covered by a countable family of codimension-two (i. e. at least two) closed sets of μ–measure zero, where μ is any smooth measure on M. (Cf. Definition 2.12 of [K-S-Sz(1991)].)

Property 2.10. The collection of all slim subsets of M is a σ-ideal, that is, countable unions of slim sets and arbitrary subsets of slim sets are also slim.

Proposition 2.11. (Locality). A subset $A \subset M$ is slim if and only if for every $x \in A$ there exists an open neighborhood U of x in M such that $U \cap A$ is slim. (Cf. Lemma 2.14 of [K-S-Sz(1991)].)

Property 2.12. A closed subset $A \subset M$ is slim if and only if $\mu(A) = 0$ and $\dim A \le \dim M - 2$.

Property 2.13. (Integrability). If $A \subset M_1 \times M_2$ is a closed subset of the product of two smooth, connected manifolds with possible boundaries and corners, and for every $x \in M_1$ the set

$$A_x = \{y \in M_2\colon\ (x, y) \in A\}$$

is slim in M_2, then A is slim in $M_1 \times M_2$.

The following propositions characterize the codimension-one and codimension-two sets.

Proposition 2.14. For any closed subset $S \subset M$ the following three conditions are equivalent:

(i) $\dim S \le \dim M - 2$;

(ii) $\mathrm{int} S = \emptyset$ and for every open connected set $G \subset M$ the difference set $G \setminus S$ is also connected;

(iii) $\mathrm{int} S = \emptyset$ and for every point $x \in M$ and for any open neighborhood V of x in M there exists a smaller open neighborhood $W \subset V$ of the point x such that for every pair of points $y, z \in W \setminus S$ there is a continuous curve γ in the set $V \setminus S$ connecting the points y and z.

(See Theorem 1.8.13 and Problem 1.8.E of [E(1978)].)

Proposition 2.15. For any subset $S \subset M$ the condition $\dim S \le \dim M - 1$ is equivalent to $\mathrm{int} S = \emptyset$. (See Theorem 1.8.10 of [E(1978)].)

We recall an elementary, but important lemma (Lemma 4.15 of [K-S-Sz(1991)]). Let R_2 be the set of phase points $x \in \mathbf{M} \setminus \partial \mathbf{M}$ such that the trajectory $S^{(-\infty,\infty)}x$ has more than one singularities (or, its only singularity is not simple).

Proposition 2.16. The set R_2 is a countable union of codimension-two smooth sub-manifolds of M and, being such, is slim.

The next lemma establishes the most important property of slim sets which gives us the fundamental geometric tool to connect the open ergodic components of billiard flows.

Proposition 2.17. If M is connected, then the complement $M \setminus A$ of a slim F_σ set $A \subset M$ is an arc-wise connected (G_δ) set of full measure. (See Property 3 of §4.1 in [K-S-Sz(1989)]. The F_σ sets are, by definition, the countable unions of closed sets, while the G_δ sets are the countable intersections of open sets.)

The subsets $\mathbf{M}^0$ and $\mathbf{M}^\#$. Denote by $\mathbf{M}^\#$ the set of all phase points $x \in \mathbf{M}$ for which the trajectory of x encounters infinitely many non-tangential collisions in both time directions. The trajectories of the points $x \in \mathbf{M} \setminus \mathbf{M}^\#$ are lines: the motion is linear and uniform, see the appendix of [Sz(1994)]. It is proven in lemmas A.2.1 and A.2.2 of [Sz(1994)] that the closed set $\mathbf{M} \setminus \mathbf{M}^\#$ is a finite union of hyperplanes. It is also proven in [Sz(1994)] that, locally, the two sides of a hyper-planar component of $\mathbf{M} \setminus \mathbf{M}^\#$ can be connected by a positively measured beam of trajectories, hence, from the point of view of ergodicity, in this paper it is enough to show that the connected components of $\mathbf{M}^\#$ entirely belong to one ergodic component. This is what we are going to do in this paper.

Denote by $\mathbf{M}^0$ the set of all phase points $x \in \mathbf{M}^\#$ the trajectory of which does not hit any singularity, and use the notation $\mathbf{M}^1$ for the set of all phase points $x \in \mathbf{M}^\#$ whose orbit contains exactly one, simple singularity. According to Proposition

2.16, the set $\mathbf{M}^{\#} \setminus (\mathbf{M}^0 \cup \mathbf{M}^1)$ is a countable union of smooth, codimension-two (≥ 2) submanifolds of $\mathbf{M}$, and, therefore, this set may be discarded in our study of ergodicity, please see also the properties of slim sets above. Thus, we will restrict our attention to the phase points $x \in \mathbf{M}^0 \cup \mathbf{M}^1$.

The "Chernov-Sinai Ansatz". An essential precondition for the Theorem on Local Ergodicity by Bálint–Chernov–Szász–Tóth (Theorem 4.4 of [B-Ch-Sz-T(2002)]) is the so called "Chernov-Sinai Ansatz" which we are going to formulate below. Denote by $\mathcal{SR}^+ \subset \partial\mathbf{M}$ the set of all phase points $x_0 = (q_0, v_0) \in \partial\mathbf{M}$ corresponding to singular reflections (a tangential or a double collision at time zero) supplied with the post-collision (outgoing) velocity v_0. It is well known that $\mathcal{SR}^+$ is a compact cell complex with dimension $2d - 3 = \dim\mathbf{M} - 2$. It is also known (see Lemma 4.1 in [K-S-Sz(1990)], in conjunction with Proposition 2.16 above) that for ν-almost every phase point $x_0 \in \mathcal{SR}^+$ the forward orbit $S^{(0,\infty)}x_0$ does not hit any further singularity. (Here ν is the Riemannian volume of $\mathcal{SR}^+$ induced by the restriction of the natural Riemannian metric of $\mathbf{M}$.) The Chernov-Sinai Ansatz postulates that for ν-almost every $x_0 \in \mathcal{SR}^+$ the forward orbit $S^{(0,\infty)}x_0$ is sufficient (geometrically hyperbolic).

The Theorem on Local Ergodicity. The Theorem on Local Ergodicity for algebraic, semi-dispersive billiards by Bálint–Chernov–Szász–Tóth (Theorem 4.4 of [B-Ch-Sz-T(2002)]) claims the following: Let $(\mathbf{M}, \{S^t\}_{t\in\mathbb{R}}, \mu)$ be a semi-dispersive billiard flow with the property that the smooth components of the boundary $\partial\mathbf{Q}$ of the configuration space are algebraic hyper-surfaces. (The cylindric billiards automatically fulfill this algebraicity condition.) Assume – further – that the Chernov-Sinai Ansatz holds true, and a phase point $x_0 \in \left(\mathbf{M}^0 \cup \mathbf{M}^1\right) \setminus \partial\mathbf{M}$ is sufficient.

Then some open neighborhood $U_0 \subset \mathbf{M}$ of x_0 belongs to a single ergodic component of the flow $(\mathbf{M}, \{S^t\}_{t\in\mathbb{R}}, \mu)$. (Modulo the zero sets, of course.)

§3. The Geometry of the Neutral Spaces Proof of Theorem 1

Suppose that $\Sigma = (\sigma_1, \sigma_2, \dots, \sigma_n)$ is a symbolic collision sequence (for N balls) that is combinatorially rich in the sense of Definition 3.28 of [Sim(2002)], i. e. for which Corollary 3.26 of that paper applies. Assume that the exceptional, closed algebraic set F (of non-sufficient phase points — despite the rich symbolic sequence Σ on their orbit segment) has a non-empty, codimension-one, smooth component. Assume, correspondingly, that an open set $U_0 \subset \mathbf{M} \setminus \partial\mathbf{M}$ and a number $T > 0$ are given with the following properties:

(1) $S^T(U_0) \cap \partial\mathbf{M} = \emptyset$,
(2) $S^{[0,T]}x$ is non-singular for all $x \in U_0$,

(3) $\Sigma\left(S^{[0,T]}x\right)=\Sigma=(\sigma_1,\sigma_2,\dots,\sigma_n)$ for all $x\in U_0$,

(4) the subset

$$\mathrm{NS}\,(\Sigma,\,U_0)=\left\{x\in U_0\big|\ \dim\mathcal{N}_0\left(S^{[0,T]}x\right)>1\right\}$$

of non-sufficient phase points is a smooth (algebraic) submanifold of U_0,

(5) the pair of sets $(U_0,\,\mathrm{NS}\,(\Sigma,\,U_0))$ is diffeomorphic to the standard pair

$$\left(\mathbb{R}^{2d-1},\,\mathbb{R}^{2d-2}\right).$$

Below we formulate (and then prove) a fundamental result, Main Lemma 3.1, about the scenario described in (1)—(5) above. In the sequel we will be frequently using the phrase "the property $P(y)$ holds true for typical (or, sometimes called "generic") points y of a smooth, algebraic manifold N". By this we will always mean that the subset

$$B=\left\{y\in N\big|\ P(y)\text{ is false}\right\}$$

is contained in a proper, closed, algebraic subset of N, i. e. B can be covered by a finite family of proper, closed, algebraic submanifolds of N. In all such cases, the exceptional set B itself will be a proper, closed, algebraic subset of the considered ambient manifold N.

The anticipated main result of this section is

Main Lemma 3.1. Use all assumptions and notations from above. We claim that the manifold $\mathrm{NS}(\Sigma,\,U_0)$ cannot coincide (even locally) with any post-singularity manifold, i. e. with any open subset of $S^\tau\left(\mathcal{SR}^+\right)$ with $\tau>0$.

Proof. The proof will be subdivided into several lemmas. The most important ingredient will be the Connecting Path Formula (CPF, Lemma 2.9 of [Sim(1992)-II], see also Proposition 2.19 in [S-Sz(1999)]). The CPF says that for the orbit segments $S^{[0,b]}x$ ($x\in U_0$, $0<b\le T$) the advance functionals $\alpha_k=\alpha(\sigma_k)$ ($k=1,\dots,m$, $\Sigma\left(S^{[0,b]}x\right)=(\sigma_1,\dots,\sigma_m)=\Sigma_m$, $0\le m\le n$) of the collisions on $S^{[0,b]}x$ are exactly the solutions of a homogeneous linear system

$$\sum_{k=1}^{m}\Gamma_{ik}\alpha_k=0,\quad i=1,2,\dots,m+P_m-N,\tag{3.2}$$

where each equation in (3.2) is a ν-dimensional vector equation, the coefficients $\Gamma_{ik}=\Gamma_{ik}\left(S^{[0,b]}x\right)\in\mathbb{R}^\nu$ of α_k are some well defined linear combinations of the relative velocities

$$(3.3)\qquad \begin{aligned}\Delta_k^- &= v_{i(k)}^{k-1} - v_{j(k)}^{k-1},\\ \Delta_k^+ &= v_{i(k)}^{k} - v_{j(k)}^{k}\end{aligned}$$

(here $\sigma_k = (i(k), j(k))$, $1 \le i(k) < j(k) \le N$, and $v_i^k \in \mathbb{R}^\nu$ is the velocity of the i-th ball between the k-th and $k+1$-st collisions), so that the coefficients of these relative velocities are given fractional linear expressions of the masses $m_{i(k)}$ and $m_{j(k)}$, please see definitions 2.17–2.18, Proposition 2.19 and formula (3.27) in [S-Sz(1999)]. Finally, $P_m = P_{\Sigma_m}$ is the number of all connected components of the collision graph of $\Sigma_m = (\sigma_1, \dots, \sigma_m)$, $0 \le m \le n$. We note that $P_n = 1$, since $\Sigma = \Sigma_n$ necessarily has a connected collision graph, thanks to the assumed combinatorial richness of Σ.

Note that, when applying the CPF, the left-hand-side of the Connecting Path Formula has to be written as the relative incoming (i. e. pre-collision) velocity of the colliding balls multiplied by the advance of that collision. It follows immediately from the exposition of [Sim(1992)-II] that the equations of this sort (arising from all CPF's) are the *only* constraints on the advance functionals. The reason why this is true is that the fulfillment of all CPF's precisely means that the relative displacement (variation of position) of every pair of particles right before their new collision is parallel to the relative incoming velocity of these particles, and this fact guarantees that the variation of the newly formed relative outgoing velocity will also be zero, just as required by the neutrality. Our understanding is that the columns of (3.2) correspond to the advances $\alpha_1, \dots, \alpha_m$, while its rows form $m + P_m - N$ ν-dimensional blocks. The important feature of (3.2) is that the dimension of the neutral space $\mathcal{N}_0\left(S^{[0,b]}x\right)$ ($x \in U_0$, $0 < b \le T$, $\Sigma\left(S^{[0,b]}x\right) = \Sigma_m = (\sigma_1, \dots, \sigma_m)$) is directly related to the rank r of (3.2):

$$(3.4)\qquad \dim \mathcal{N}_0\left(S^{[0,b]}x\right) = \nu(P_m - 1) + m - r,$$

see Proposition 3.4 in [Sim(1992)-II]. (For $m = 0$, i. e. when $S^{[0,b]}x$ has no collision at all, the neutral space is – by definition – the d-dimensional space of all spatial variations δq of $x = (q, v)$ at time $t = 0$. Of course, in that case the vector space $\mathcal{N}_0\left(S^{[0,b]}x\right)$ is naturally identified with the tangent space of the parallelizable configuration space $\mathbf{Q}$, $d = \dim\mathbf{Q} = \nu(N-1)$, $P_0 = N$, $r = 0$, $m = 0$.) We note that the term $m - r$ in (3.4) is the dimension of the solution set $\{\alpha_1, \dots, \alpha_m\}$ of (3.2) which, according to Proposition 3.4 of [Sim(1992)-II], is the rank of the mapping $\Phi_m : \mathcal{N}_0\left(S^{[0,b]}x\right) \to \mathbb{R}^m$, $\Phi_m(W) = (\alpha_1(W), \dots, \alpha_m(W))$. As a special, though important, case of (3.4), we mention that if $P_m = 1$ (i. e. the collision graph of $S^{[0,b]}x$ is connected), then $S^{[0,b]}x$ is sufficient if and only if $r = m-1$ or, equivalently, if the only solution of (3.2) is the always existing trivial one corresponding to the flow direction: $\alpha_1 = \alpha_2 = \cdots = \alpha_m$, see also Remark 4.5 in [S-Sz(1999)]. It follows from (3.4) and what has been said before that for any given positive integer D, for any $x \in U_0$, and for any b $(0 < b \le T)$ the inequality

$$\dim \mathcal{N}_0 \left(S^{[0,b]} x \right) > D$$

is true if and only if the determinant $\det M$ vanishes for every square shaped submatrix M of (3.2) with the size

$$[\nu(P_m - 1) + m - D] \times [\nu(P_m - 1) + m - D] \, .$$

In other words, the "event" $\dim \mathcal{N}_0 \left(S^{[0,b]} x \right) > D$ is equivalent to the simultaneous vanishing of finitely many smooth algebraic functions $\det M$. Since such a smooth (analytic) algebraic function $\det M$ is either identically zero on U_0 or it only vanishes on a proper, closed, algebraic subset F of U_0, we proved

Lemma 3.5. For every integer m $(0 \le m \le n)$ there are two typical dimensions $D_1(m)$ and $D_2(m)$ such that

$$\dim \mathcal{N}_0 \left(S^{[0,b_m]} x \right) = D_1(m)$$

for typical phase points $x \in U_0$, and

$$\dim \mathcal{N}_0 \left(S^{[0,b_m]} x \right) = D_2(m)$$

for typical phase points $x \in \mathrm{NS}(\Sigma, U_0)$. Here the threshold b_m $(0 < b_m \le T)$ denotes a selected (and then fixed) moment of time, such that $S^{b_m}(U_0) \cap \partial \mathbf{M} = \emptyset$ and $\Sigma \left(S^{[0,b_m]} x \right) = \Sigma_m = (\sigma_1, \dots, \sigma_m)$ for all $x \in U_0$.

Note that the obvious relations

$$\begin{aligned} D_1(m) &= \min_{x \in U_0} \dim \mathcal{N}_0 \left(S^{[0,b_m]} x \right), \\ D_2(m) &= \min_{x \in \mathrm{NS}(\Sigma,\, U_0)} \dim \mathcal{N}_0 \left(S^{[0,b_m]} x \right) \end{aligned} \tag{3.6}$$

hold and, consequently, we have

$$D_1(m) \le D_2(m), \tag{3.7}$$

$m = 0, 1, \dots, n$.

Remark 3.8. It follows immediately from the assumptions (1)–(5) (preceding Main Lemma 3.1) and from Lemma 3.5 that $1 = D_1(n) < D_2(n)$. The inequalities $D_i(m_1) \geq D_i(m_2)$ are obvious for $0 \leq m_1 < m_2 \leq n$, $i = 1, 2$.

By switching from the open set U_0 to a smaller open set $\tilde{U}_0 \subset U_0$, if necessary, we can assume that for the pair of sets $\left(\tilde{U}_0,\, \mathrm{NS}\left(\tilde{U}_0,\, \Sigma\right)\right)$ the properties (1)–(5) from the beginning of this section are still true and, in addition,

$$\begin{aligned} D_1(m) &= \dim \mathcal{N}_0\left(S^{[0,b_m]}x\right), \\ D_2(m) &= \dim \mathcal{N}_0\left(S^{[0,b_m]}y\right) \end{aligned} \tag{3.9}$$

for every $x \in \tilde{U}_0 \setminus \mathrm{NS}\left(\tilde{U}_0,\, \Sigma\right)$, $y \in \mathrm{NS}\left(\tilde{U}_0,\, \Sigma\right)$, $0 \leq m \leq n$. In the future the new pair $\left(\tilde{U}_0,\, \mathrm{NS}\left(\tilde{U}_0,\, \Sigma\right)\right)$ will also be denoted (a bit sloppily) by $(U_0,\, \mathrm{NS}\,(U_0,\, \Sigma))$.

Definition. Any system of homogeneous linear equations

$$\sum_{k=1}^{m} \Gamma_{ik}\alpha_k = 0 \ (\in \mathbb{R}^\nu), \quad i = 1, \ldots, s \tag{3.2/a}$$

will be called a *permitted system* if each coefficient $\Gamma_{ik} \in \mathbb{R}^\nu$ is a given linear combination of the relative velocities Δ_k^- and Δ_k^+ (see (3.3)), such that the coefficients of these relative velocities are given rational expressions of degree zero of the masses $m_{i(k)}$ and $m_{j(k)}$ participating in the k-th collision $\sigma_k = (i(k),\, j(k))$. A determinant $\det M$ will be called a *permitted determinant* if M is a square shaped submatrix of the coefficient matrix of a permitted system (3.2/a).

We have seen that the smooth algebraic submanifold $\mathrm{NS}\,(U_0,\, \Sigma)$ of U_0 can be defined by a single equation

$$\mathrm{NS}\,(U_0,\, \Sigma) = \left\{x \in U_0 \,\middle|\, \det M_0(x) = 0\right\} \tag{3.10}$$

with a permitted determinant $\det M_0 = \det M_0(x)$. (Once a finite family of such determinant equations defines a smooth, codimension-one submanifold, clearly, one of those equations already does the job.)

Select a defining permitted determinant equation $\det M_0(x) = 0$ of (3.10) with the property that the largest column index k_0 of M_0 is the minimal possible one out of all such column indices. The index k_0 will be called the *critical index*, and the corresponding collision σ_{k_0} the *critical collision*.

In the future we will need

Lemma 3.11. For any phase point $x \in U_0$ the following two assertions are equivalent:

(1) There exists a point $y \in U_0$ such that

$$\frac{\Delta_k^-(y)}{||\Delta_k^-(y)||} = \frac{\Delta_k^-(x)}{||\Delta_k^-(x)||},\ \frac{\Delta_k^+(y)}{||\Delta_k^+(y)||} = \frac{\Delta_k^+(x)}{||\Delta_k^+(x)||} \text{ for } k = 1, \dots, k_0 - 1,$$

but at least one of the two equations

$$\frac{\Delta_{k_0}^-(y)}{||\Delta_{k_0}^-(y)||} = \frac{\Delta_{k_0}^-(x)}{||\Delta_{k_0}^-(x)||},\ \frac{\Delta_{k_0}^+(y)}{||\Delta_{k_0}^+(y)||} = \frac{\Delta_{k_0}^+(x)}{||\Delta_{k_0}^+(x)||}$$

is false;

(2) $\mathcal{N}_0\left(S^{[0,b_{k_0-1}]}x\right) \neq \mathcal{N}_0\left(S^{[0,b_{k_0}]}x\right)$.

Proof. One easily sees that for fixed directions

$$\frac{\Delta_k^\pm(x)}{||\Delta_k^\pm(x)||},\ k = 1, \dots, k_0 - 1$$

the permitted perturbations $(\delta q_1, \dots, \delta q_N; \delta v_1, \dots, \delta v_N)$ of the phase point $x = (q, v) \in U_0$ (not changing the above unit vectors) are exactly the ones for which $(\delta q_1, \dots, \delta q_N) \in \mathcal{N}_0\left(S^{[0,b_{k_0-1}]}x\right)$, and $(\delta v_1, \dots, \delta v_N) \in \mathcal{N}_0\left(S^{[0,b_{k_0-1}]}x\right)$. Therefore, if (2) is false, then for any perturbed point $y = (q + \delta q;\, v + \delta v)$ of $x = (q, v) = (\delta q_1, \dots, \delta q_N; \delta v_1, \dots, \delta v_N)$ (for which the equations

$$\frac{\Delta_k^-(y)}{||\Delta_k^-(y)||} = \frac{\Delta_k^-(x)}{||\Delta_k^-(x)||},\ \frac{\Delta_k^+(y)}{||\Delta_k^+(y)||} = \frac{\Delta_k^+(x)}{||\Delta_k^+(x)||} \text{ for } k = 1, \dots, k_0 - 1$$

of (1) hold true) the difference vectors $\delta q = (\delta q_1, \dots, \delta q_N)$, and $\delta v = (\delta v_1, \dots, \delta v_N)$ belong to $\mathcal{N}_0\left(S^{[0,b_{k_0}]}x\right)$, thus

$$\frac{\Delta_{k_0}^\pm(y)}{||\Delta_{k_0}^\pm(y)||} = \frac{\Delta_{k_0}^\pm(x)}{||\Delta_{k_0}^\pm(x)||},$$

i. e. (1) is also false.

On the other hand, if (2) is true, then one chooses an element

$$\delta q = (\delta q_1, \dots, \delta q_N) \in \mathcal{N}_0\left(S^{[0,b_{k_0-1}]}x\right) \setminus \mathcal{N}_0\left(S^{[0,b_{k_0}]}x\right)$$

(with a $||\delta q||$ small enough), and forms the perturbed phase point $y = (q + \delta q,\, v)$ for the original point $x = (q, v) \in U_0$. It follows immediately from the relation

$\delta q \notin \mathcal{N}_0\left(S^{[0,b_{k_0}]}x\right)$ that $\Delta^+_{k_0}(y)$ is not parallel to $\Delta^+_{k_0}(x)$, thus finishing the proof of the lemma. $\square$

The minimality of the critical column index k_0 and the fact that $\mathrm{NS}\,(\Sigma,\, U_0) \neq U_0$ imply that

(i) the permitted determinant $\det M_0$ does depend on its last column (of index k_0), i. e. for any given collection $\Delta^\pm_k$, $k = 1, 2, \dots, k_0 - 1$ of relative velocities, the equation $\det M_0 = 0$ defines a proper (i. e. $(2\nu - 1)$-dimensional) linear subspace in the space of velocities $\left(\Delta^-_{k_0},\, \Delta^+_{k_0}\right)$;

(ii) in Lemma 3.11 the two equivalent assertions (1) and (2) are actually true for typical phase points $x \in \mathrm{NS}\,(\Sigma,\, U_0)$. Otherwise the submanifold $\mathrm{NS}\,(\Sigma,\, U_0)$ (or, in other words, the validity of the equation $\det M_0 = 0$) would be determined by the velocity history $\Delta^\pm_k$, $k = 1, 2, \dots, k_0 - 1$, thus violating the minimality of the index k_0.

We may assume, as before, that (1) and (2) of 3.11 are true not only for typical phase points $x \in \mathrm{NS}\,(\Sigma,\, U_0)$, but for all points $x \in \mathrm{NS}\,(\Sigma,\, U_0)$.

Construction of the Transversal Curves.

For the points $x = (q, v)$ of $\mathrm{NS}\,(\Sigma,\, U_0)$ select the vectors

$$\delta q = \delta q(x) = (\delta q_1, \dots, \delta q_N) \in \mathcal{N}_0\left(S^{[0,b_{k_0-1}]}x\right) \setminus \mathcal{N}_0\left(S^{[0,b_{k_0}]}x\right),$$

depending analytically on x, and choose the positive numbers $\kappa = \kappa(x)$, which will play the role of τ^{-1} with the number $\tau = \tau(x)$ in Main Lemma 3.1. Assign to the phase point x the curve

$$\gamma = \gamma_x = \left\{\pi_t(x) = (q + t\delta q,\, v + t\kappa\delta q)\big|\ |t| < \epsilon_0\right\} \tag{3.12}$$

wity a suitably small $\epsilon_0 > 0$, selected uniformly for all $x \in \mathrm{NS}\,(\Sigma,\, U_0)$.

Remark 3.13. The reader might get confused by the fact that the linear perturbation $v + t\kappa\delta q$ of the velocity leads us out of the phase space (constituted by phase points with a unit speed). This, however, should cause no problem in the proof: The entire formalism is invariant under rescaling the velocities. If somebody feels uncomfortably with this, he/she can rescale the velocity of $\pi_t(x)$ to get a unit velocity vector.

An important feature of the perturbation $\pi_t(x)$ of x is that

$$\frac{\Delta^\pm_k(\pi_t(x))}{\big|\big|\Delta^\pm_k(\pi_t(x))\big|\big|} = \frac{\Delta^\pm_k(x)}{\big|\big|\Delta^\pm_k(x)\big|\big|}, \quad k = 1, 2, \dots, k_0 - 1, \tag{3.14}$$

thanks to the relation $\delta q \in \mathcal{N}_0\left(S^{[0,b_{k_0}-1]}x\right)$. According to the property (i) above, for the given values of $\Delta_k^{\pm}(x)$, $k=1,2,\dots,k_0-1$, the equation $\det M_0(x)=0$ defines a $(2\nu-1)$-dimensional subspace

$$H_x=\left\{(\Delta_{k_0}^-,\,\Delta_{k_0}^+)\in\mathbb{R}^\nu\oplus\mathbb{R}^\nu\,\middle|\,\langle\Delta_{k_0}^-,\,A\rangle+\langle\Delta_{k_0}^+,\,B\rangle=0\right\} \tag{3.15}$$

in the space of pairs of velocities $(\Delta_{k_0}^-,\,\Delta_{k_0}^+)\in\mathbb{R}^\nu\oplus\mathbb{R}^\nu=\mathbb{R}^{2\nu}$. Here the vectors $A=A(x)\in\mathbb{R}^\nu$, $B=B(x)\in\mathbb{R}^\nu$ depend on $x\in U_0$. Plainly, we can assume that this dependence is analytic, and $A(x)=A(y)$, whenever $\mathcal{N}_0\left(S^{[0,b_{k_0}-1]}x\right)=\mathcal{N}_0\left(S^{[0,b_{k_0}-1]}y\right)$. (The angled bracket denotes the standard inner product in $\mathbb{R}^\nu$.) We observe first that, due to (3.14) above, the same vectors A, $B\in\mathbb{R}^\nu$ define the $(2\nu-1)$-dimensional subspace $H_{\pi_t(x)}$ associated with the perturbed point $\pi_t(x)$, $|t|<\epsilon_0$. The next step in the proof of Main Lemma 3.1 is

Lemma 3.16. For typical $x\in\mathrm{NS}\,(\Sigma,\,U_0)$ and $t\in(-\epsilon_0,\,\epsilon_0)$ the vector

$$\left(\Delta_{k_0}^-(\pi_t(x)),\,\Delta_{k_0}^+(\pi_t(x))\right)$$

does not belong to $H_x=H_{\pi_t(x)}$, i. e.

$$\left\langle\Delta_{k_0}^-(\pi_t(x)),\,A\right\rangle+\left\langle\Delta_{k_0}^+(\pi_t(x)),\,B\right\rangle\neq0.$$

Proof. The function

$$\phi(t)=\left\langle\Delta_{k_0}^-(\pi_t(x)),\,A\right\rangle+\left\langle\Delta_{k_0}^+(\pi_t(x)),\,B\right\rangle$$

is clearly an analytic algebraic function of t, $|t|<\epsilon_0$. Therefore, the only alternative to an almost everywhere nonzero function ϕ is that $\phi(t)\equiv0$. By using a fixed number (a time moment) s_0 slightly smaller than $t_{k_0}(x)$ (the time of σ_{k_0} on the orbit of x), we express $S^{s_0}\left(\pi_t(x)\right)$ in local coordinates as a function of t:

$$S^{s_0}\left(\pi_t(x)\right)=S^{s_0}(x)+\left((\kappa^{-1}+s_0)t\kappa\delta\tilde{q},\,t\kappa\delta\tilde{q}\right), \tag{3.17}$$

where $\delta\tilde{q}=R_{k_0-1}R_{k_0-2}\dots R_1\delta q$, and R_i is the orthogonal velocity reflection at the boundary $\partial\mathbf{Q}$ at the point $q\left(S^{t_i(x)}(x)\right)$. Here we used the definition (3.12) of $\pi_t(x)$, and the neutrality of $\delta q=\delta q(x)=(\delta q_1,\dots,\delta q_N)$ with respect to $\sigma_1,\dots,\sigma_{k_0-1}$. If we switch from (3.17) to the coordinates describing merely the relative motion of the $i(k_0)$-th and $j(k_0)$-th balls (i. e. we only consider $q_{i(k_0)}-q_{j(k_0)}$ and $v_{i(k_0)}-v_{j(k_0)}$), then, in order to prove that $\phi(t)\not\equiv0$, it is enough to prove

Sub-lemma 3.18. Let A, B be vectors in $\mathbb{R}^\nu$ (so that at least one of them is non-zero), $C = C_{2r} \subset \mathbb{R}^\nu$ be the sphere of radius $2r$ in $\mathbb{R}^\nu$ centered at the origin, $Q \in \mathbb{R}^\nu$ a point lying outside of C, $0 \neq V = \tilde{v}_{i(k_0)} - \tilde{v}_{j(k_0)} \in \mathbb{R}^\nu$ a (velocity) vector such that the line $l \subset \mathbb{R}^\nu$ through Q in the direction of V intersects C in two distinct points $P = P_1$ and P_2 (the point P being closer to Q, V pointing in the direction of $\overrightarrow{QP}$), $d = \kappa^{-1} + s_0$ be the distance between Q and P, and let $w = \kappa(\delta\tilde{q}_{i(k_0)} - \delta\tilde{q}_{j(k_0)}) \in \mathbb{R}^\nu$ be a vector not parallel to V. For any t, $|t| < \epsilon_0$, let $V_t = V + tw$, $l_t \subset \mathbb{R}^\nu$ be the line through Q in the direction of V_t, P_t the point of $l_t \cap C$ closer to Q, and, finally, V_t^+ the reflection of V_t across the tangent hyperplane of C at P_t. We claim that for typical phase points $x = (q,\, v) \in \mathrm{NS}\,(\Sigma,\, U_0)$ (featuring (3.17)) and for almost every selection of $\delta q = \delta q(x) = (\delta q_1, \dots, \delta q_N) \in \mathcal{N}_0\left(S^{[0,b_{k_0}-1]}x\right) \setminus \mathcal{N}_0\left(S^{[0,b_{k_0}]}x\right)$ the smooth (analytic) algebraic function

$$\phi(t) = \langle V_t,\, A\rangle + \langle V_t^+,\, B\rangle$$

is not identically zero.

Proof. Assume that $\phi(t) \equiv 0$. Let $t_1 < 0 < t_2$ be the two values of $t \in \mathbb{R}$ for which the lines l_{t_1} and l_{t_2} are tangent to the sphere C. Since $V_{t_i}^+ = V_{t_i}$ $(i = 1,\, 2)$, according to our indirect assumption $\langle V_{t_i},\, A + B\rangle = 0$ for $i = 1, 2$, thus the vector $A + B \in \mathbb{R}^\nu$ is orthogonal to the $2D$ plane $\Pi = \Pi(x)$ spanned by $V_0 = V = \tilde{v}_{i(k_0)} - \tilde{v}_{j(k_0)}$ and $w = \kappa(\delta\tilde{q}_{i(k_0)} - \delta\tilde{q}_{j(k_0)})$. By writing $A = -B + X$, $X \perp \Pi$, the assumed identity $\phi(t) \equiv 0$ means that $\langle V_t,\, -B + X\rangle + \langle V_t^+,\, B\rangle \equiv 0$, i. e.

$$(*) \qquad \langle V_t^+ - V_t,\, B\rangle \equiv 0.$$

If the translated 2-plane $Q + \Pi \subset \mathbb{R}^\nu$ does not contain the origin, then, as it is easy to see, the arising difference vectors $V_t^+ - V_t$ span the entire space $\mathbb{R}^\nu$, which means that $B = 0$. So we got that either $B = B(x) = 0$ for every phase point $x = (q,\, v) \in \mathrm{NS}\,(\Sigma,\, U_0)$, or $Q \in \Pi$ for all points $x = (q,\, v) \in \mathrm{NS}\,(\Sigma,\, U_0)$ and every selection of $\delta q = \delta q(x) = (\delta q_1, \dots, \delta q_N) \in \mathcal{N}_0\left(S^{[0,b_{k_0}-1]}x\right) \setminus \mathcal{N}_0\left(S^{[0,b_{k_0}]}x\right)$. In the first case, due to the assumtion $\phi(t) \equiv 0$ for all $x \in \mathrm{NS}\,(\Sigma,\, U_0)$ and for every selection of $\delta q = \delta q(x) = (\delta q_1, \dots, \delta q_N) \in \mathcal{N}_0\left(S^{[0,b_{k_0}-1]}x\right) \setminus \mathcal{N}_0\left(S^{[0,b_{k_0}]}x\right)$, we get that the non-zero vector $A = A(x)$ is always orthogonal to Π, which means that the validity of $\det M_0(x) = 0$ $(x \in U_0)$ is already determined by the unit vectors

$$\frac{\Delta_k^\pm(x)}{||\Delta_k^\pm(x)||}, \quad k = 1, \dots, k_0 - 1,$$

thus contradicting to the minimality of the critical index k_0. In the second case, however, the equation (*) above implies that $B = B(x) \perp \Pi$, so $A = B + X \perp \Pi$, as well, and again the validity of $\det M_0(x) = 0$ $(x \in U_0)$ would be determined by the unit vectors

$$\frac{\Delta_k^{\pm}(x)}{||\Delta_k^{\pm}(x)||}, \quad k = 1, \dots, k_0 - 1,$$

contradicting to the minimality of k_0. This finishes the indirect proof of the sub-lemma. $\square$

The successful proof of the sub-lemma completes the proof of Lemma 3.16, as well. $\square$

The last lemma in the proof of Main Lemma 3.1 is

Lemma 3.19. For typical phase points $x \in \mathrm{NS}\,(\Sigma,\, U_0)$ the curve γ_x (defined in (3.12)) is transversal to the manifold $\mathrm{NS}\,(\Sigma,\, U_0)$.

Proof. Observe, first of all, that the direction vector $\delta q = \delta q(x) \in \mathcal{N}_0\left(S^{[0,b_{k_0}-1]}x\right) \setminus \mathcal{N}_0\left(S^{[0,b_{k_0}]}x\right)$ (defining γ_x via (3.12)) can be selected in such a way that $\delta q(x_1) = \delta q(x_2)$ for any pair of points $x_1,\, x_2 \in \mathrm{NS}\,(\Sigma,\, U_0)$ sharing the same neutral spaces $\mathcal{N}_0\left(S^{[0,b_{k_0}]}x\right)$ and $\mathcal{N}_0\left(S^{[0,b_{k_0}-1]}x\right)$. Then the angle $\psi(x)$ made by γ_x and $\mathrm{NS}\,(\Sigma,\, U_0)$ at x is also an analytic function of x. The function $\psi(x)$ cannot be identically zero on $\mathrm{NS}\,(\Sigma,\, U_0)$, for in that case we would have $\gamma_x \subset \mathrm{NS}\,(\Sigma,\, U_0)$, $\phi_x(t) \equiv 0$ for each $x \in \mathrm{NS}\,(\Sigma,\, U_0)$, violating Lemma 3.16. Thus $\psi(x) > 0$ for typical $x \in \mathrm{NS}\,(\Sigma,\, U_0)$. $\square$

Finishing the proof of Main Lemma 3.1. Let $x \in S^{\tau}\left(\mathcal{SR}^+\right) \cap \mathrm{NS}\,(\Sigma,\, U_0)$. The proof in the paragraph of [K-S-Sz(1990)] containing the Sub-lemma 4.4 shows that the curve γ_x of (3.12) is tangential to the manifold $S^{\tau}\left(\mathcal{SR}^+\right)$, yet it is transversal to $\mathrm{NS}\,(\Sigma,\, U_0)$ for typical $x \in \mathrm{NS}\,(\Sigma,\, U_0)$. This proves that the codimension-one submanifolds $S^{\tau}\left(\mathcal{SR}^+\right)$ and $\mathrm{NS}\,(\Sigma,\, U_0)$ cannot even locally coincide, thereby finishing the proof of Main Lemma 3.1. $\square$

Corollary: The Proof of Theorem 1.

As it is explained in 9.3 of [Sim(2003)], it is exclusively the proof of Proposition 3.1 (of that paper) that uses the rather involved algebraic machinery which, in turn, makes it necessary to drop an unspecified zero-measured set of the external geometric parameters $(m_1, \dots, m_N, r)$ to prove the (hyperbolic) ergodicity of the standard hard disk system. Our Main Lemma 3.1 above, however, proves exactly the statement of 3.1 of [Sim(2003)] for each hard disk system, without exceptional $(N+1)$-tuples of external geometric parameters $(m_1, \dots, m_N, r)$. Therefore, the results of [Sim(2003)], together with Main Lemma 3.1 of the present paper, prove our Theorem 1, i. e. that every standard hard disk system ($\nu = 2$) is (fully hyperbolic and) ergodic. $\square$

§4. What Can Be Saved in Higher Dimensions? Proof of Theorem 2

Assume that $\nu \geq 3$, $N \geq 2$, $m_1, \dots, m_N > 0$, $r > 0$, and consider the standard toroidal hard ball system $\left(\mathbf{M}_{\vec{m},r}, \left\{S^t_{\vec{m},r}\right\}, \mu_{\vec{m},r}\right) = (\mathbf{M}, \{S^t\}_{t\in\mathbb{R}}, \mu)$ with a connected interior of the phase space $\mathbf{M}$ (as always). The Theorem in [Sim(2002)] (more precisely, its corollary presented right afterwards) claims that every such system $(\mathbf{M}, \{S^t\}_{t\in\mathbb{R}}, \mu)$ is fully hyperbolic (i. e. has nonzero relevant Lyapunov exponents almost everywhere), has ergodic components C_i of positive measure, and the restriction $\left\{S^t \middle| C_i\right\}_{t\in\mathbb{R}}$ of the flow to any component C_i has the Bernoulli mixing property. By applying our current Main Lemma 3.1 in the case $\nu \geq 3$, we will prove the following theorem, which is actually a bit stronger than Theorem 2 (see §1) by also verifying the Chernov–Sinai Ansatz:

Theorem 2/a. For every toroidal hard ball system $\left(\mathbf{M}_{\vec{m},r}, \left\{S^t_{\vec{m},r}\right\}, \mu_{\vec{m},r}\right)$ ($N \geq 2$, $\nu \geq 3$)

(1) the Chernov–Sinai Ansatz (see §2) holds true, and
(2) the ergodic components $C_i \subset \mathbf{M}$ of the flow $\{S^t\}_{t\in\mathbb{R}}$ are open subsets of $\mathbf{M}$ (modulo the zero measured sets).

We remind the reader that (2) is, in fact, a corollary of (1), the Theorem on Local Ergodicity for algebraic semi-dispersive billiards (Theorem 4.4 of [B-Ch-Sz-T(2002)]), and the full hyperbolicity of the considered standard hard ball flow (i. e. the theorem of [Sim(2002)]). We also point out that the essence of Theorem 2/a is that it holds for *every* hard ball system (without exceptional $(N+1)$-tuples $(m_1, \dots, m_N, r)$ of the external geometric parameters), unlike the ergodicity result of [Sim(2004)-I].

Proof of Theorem 2/a by induction on N. In the case $N = 2$ the statements have been known to be true since [S-Ch(1987)], for in this case every trajectory (with at least one proper collision) is automatically sufficient (geometrically hyperbolic), and the system is actually ergodic.

Assume now that $N \geq 3$, and Theorem 2/a has been proved for all situations with N' balls, $2 \leq N' < N$. We want to prove (1) for any system $\left(\mathbf{M}_{\vec{m},r}, \left\{S^t_{\vec{m},r}\right\}, \mu_{\vec{m},r}\right)$ with N particles.

Consider, therefore, a smooth component (an open cell) $C \subset \mathcal{SR}^+$ of $\mathcal{SR}^+$ with the maximum dimension $2d-3$, and let $\tilde{\nu} = \tilde{\nu}_C$ be the measure on C induced by the restriction of the Riemannian metric of $\mathbf{M}$ to C. We need to prove that for $\tilde{\nu}$-almost every phase point $x \in C$ the forward orbit $S^{(0,\infty)}x$ is non-singular and hyperbolic.

The non-singularity of $S^{(0,\infty)}x$ (for $\tilde{\nu}$-almost every $x \in C$) is a direct consequence of Lemma 4.1 of [K-S-Sz(1990)]. It will turn out that, in order to prove that $S^{(0,\infty)}x$ is hyperbolic for typical points $x \in C$, it is enough to prove

Main Lemma 4.1. Use all of the above assumptions and notations, including the induction hypothesis of Theorem 2/a. We claim that for every partition π: $\{1, 2, \dots, N\} = P_1 \cup \dots \cup P_k$, $k \geq 2$, of the label set $\{1, 2, \dots, N\}$ and for every threshold $T > 0$ the subset

$$F = F_{C,\pi,T} = \big\{ x \in C \big| \text{ on } S^{[T,\infty)}x \text{ there is no proper collision between particles from different classes of } \pi \big\} \tag{4.2}$$

of C has $\tilde{\nu}$-measure zero.

Proof. Since the assertion of the main lemma is a local statement, we can switch from C to any open subset $C' \neq \emptyset$ of it, and assume that $S^T(C') \cap \partial\mathbf{M} = \emptyset$. In the future the set C' will be denoted by C, a bit sloppily.

We will be studying small perturbations $\delta y = (\delta q,\, \delta v) \in \mathcal{T}_y\mathbf{M}$ of the points $y = (q,\, v) = S^T x$ $(x \in C)$ and, correspondingly, some orthogonal direct sum decompositions of the tangent spaces $\mathcal{T}_y\mathbf{M}$ of $\mathbf{M}$ at points $y \in S^T(C)$, so that these decompositions will “reflect” the non-interaction between particles belonging to different classes of π.

By performing an easy induction on the number $N - k$ (where $k = |\pi|$), we see that it suffices to prove that

$$\tilde{\nu}\left(F_{C,\pi,T} \setminus \bigcup_{\substack{\pi',\, T' \\ \pi' > \pi \\ T' \geq T}} F_{C,\pi',T'} \right) = 0, \tag{4.3}$$

where $\pi' > \pi$ indicates that π' is a proper refinement of the partition π. (Of course, in the case $k = N$ the union in (4.3) is empty.) According to Lemma 4.1 of [K-S-Sz(1990)], the doubly singular phase points form a countable family of codimension-two submanifolds, thus typical phase points x of C do not encounter any further singularity on their forward orbit $S^{(0,\infty)}x$. Therefore, it is enough to study the points y in the subset

$$\tilde{F} = \left\{ y = S^T x \big| \; x \in F_{C,\pi,T} \setminus \bigcup_{\substack{\pi',\, T' \\ \pi' > \pi \\ T' \geq T}} F_{C,\pi',T'}, \quad S^{(0,\infty)}x \text{ is non-singular} \right\} \tag{4.4}$$

of $S^T(C)$. For the points $y \in \tilde{F}$ we introduce the following linear subspaces of $\mathcal{T}_y\mathbf{M}$:

$$\begin{aligned}(4.5)\qquad & W_i = W_i(y) = \Big\{\delta y = (\delta q,\, \delta v) \in \mathcal{T}_y\mathbf{M} \Big| \; \delta q_j = \delta v_j = 0 \text{ for } j \notin P_i, \\ & \sum_{j\in P_i} m_j \delta q_j = \sum_{j\in P_i} m_j \delta v_j = 0,\; \sum_{j\in P_i} m_j \langle v_j,\, \delta q_j\rangle = \sum_{j\in P_i} m_j \langle v_j,\, \delta v_j\rangle = 0, \\ & (\delta v_j)_{j\in P_i} = -B_i(y)\cdot (\delta q_j)_{j\in P_i} \Big\},\end{aligned}$$

where the non-negative operator $B_i = B_i(y)$ is defined for the internal subsystem P_i by (2.4) of [K-S-Sz(1990)] (the P_i-part

$$y_i = \Big((\delta q_j)_{j\in P_i},\, (\delta v_j)_{j\in P_i} \Big)$$

of y plays the role of x in (2.4)) as the negative of the second fundamental form of the exponentially stable manifold $\gamma^{(s)}_{\exp}(y_i)$ of the P_i subsystem. The P_i-subdynamics $S_i^{(0,\infty)} y_i$ of y_i is a combinatorially rich forward orbit (since

$$S^{-T}y = x \notin \bigcup_{\substack{\pi',\, T' \\ \pi' > \pi \\ T' \ge T}} F_{C,\pi',T'},$$

yet $S_i^{(0,\infty)} y_i$ still may be nonsufficient on a codimension-one, exceptional submanifold E. However, the tangent plane $\mathcal{T}_y S^T(C)$ of $S^T(C)$ at y contains the umbilic subspace $\big\{(\delta q,\, \delta v) \big| \; \delta v = \kappa\delta q\big\}$ (where $\kappa = T^{-1}$) so, according to our Main Lemma 3.1 above, the manifolds $S^T(C)$ and E cannot locally coincide, i. e. $\operatorname{codim}\big(S^T(C)\cap E\big) \ge 2$. Thus we have proved

Lemma 4.6. For $\tilde{\nu}$-almost every point $y \in C' = S^T(C)$ the non-negative operators B_i in (4.5) are actually positive definite for $i = 1, \dots, k$, that is, the sub-billiard forward orbits $S_i^{(0,\infty)} y_i$ are sufficient (geometrically hyperbolic). $\square$

Remark. In the case $|P_i| = 1$ the linear space W_i is the null space, and the zero operator B_i (also acting on a null space) is considered to be positive definite, in accordance with the usual convention that the trivial, one-particle forward orbit $S_i^{(0,\infty)} y_i$ is considered to be sufficient.

Next we define the "special neutral space" N_π as follows:

$$\begin{aligned}(4.8)\qquad & N_\pi = N_\pi(y) = \Big\{\delta y = (\delta q,\, \delta v) \in \mathcal{T}_y\mathbf{M} \Big| \; \delta v = 0, \quad \delta q_{j_1} - \delta q_{j_2} = \tau_i(v_{j_1} - v_{j_2}) \\ & \text{for any } j_1,\, j_2 \in P_i \text{ with some } \tau_i \in \mathbb{R}, \quad i = 1, \dots, k\Big\}.\end{aligned}$$

It is clear that the linear subspaces $W_1, \dots, W_k$ and N_π are mutually orthogonal, $\dim W_i = |(|P_i| - 1)\nu - 1|_+$ (in the case $|P_i| = 1$ we have $\dim W_i = 0$), and

$$\dim N_\pi = (k-1)\nu + \#\left\{i \,\middle|\, |P_i| > 1\right\},$$

thus the orthogonal direct sum

$$W = W(y) = N_\pi \oplus \bigoplus_{i=1}^{k} W_i \tag{4.9}$$

has the dimension $(N-1)\nu = d = \dim\mathbf{Q}$. Clearly, the space W projects isometrically onto the tangent space of $\mathbf{Q}$ by the projection $(\delta q,\, \delta v) \mapsto \delta q$, and W actually assumes the form

$$W = \left\{(\delta q,\, -B\delta q) \,\middle|\, \delta q \in \mathcal{T}\mathbf{Q}\right\}, \tag{4.10}$$

where the operator $B \geq 0$ is the direct sum of all the positive operators B_i ($i = 1, \dots, k$) and the zero operator corresponding to the space N_π. Our next result is a fundamental transversality statement:

Lemma 4.11. Assume that the phase point $y \in \tilde{F}$ (for the definition of $\tilde{F}$, please see (4.4)) all operators $B_i = B_i(y_i)$ are positive definite. Then the space $W = W(y)$ of (4.9) is *transversal* to the flow-invariant hull

$$J = \bigcup_{t>0} S^t(C)$$

of C at the point y.

Proof. Consider a normal vector $0 \neq n_0 = (z,\, w) \in \mathcal{T}_y\mathbf{M}$ of the codimension-one submanifold J at the point y. According to (7.11) of [Sim(2003)] (or, by the results of §3 of [Sim(2004)-II], esp. (3.12)) we have the inequality $\langle z,\, w\rangle < 0$. Since the operator B of (4.10) is non-negative, we have that $Bw \neq z$. Select a vector $\delta q \in \mathcal{T}\mathbf{Q}$ with the property $\langle \delta q,\, z - Bw\rangle \neq 0$. Then the element $(\delta q,\, -B\delta q)$ of W is not perpendicular to $n_0 = (z,\, w)$:

$$\langle(\delta q,\, -B\delta q),\, (z,\, w)\rangle = \langle \delta q,\, z\rangle - \langle \delta q,\, Bw\rangle = \langle \delta q,\, z - Bw\rangle \neq 0.$$

This finishes the proof of Lemma 4.11. $\square$

So far we got that either some of the subspaces $W_i = W_i(y)$ is transversal to J, or N_π has this property. We will finish the proof of Main Lemma 4.1 by way of contradiction. Thus, assume that $\tilde{\nu}(\tilde{F}_+) > 0$, where

$$\tilde{F}_+ = \left\{y \in \tilde{F} \,\middle|\, B_i(y_i) > 0, \quad i = 1, \dots, k\right\}. \tag{4.12}$$

By the transversality property explained above, either there exists an index i, for instance $i=1$, such that $W_1(y)$ is transversal to J at y for a $\tilde{\nu}$-positive set of points $y\in\tilde{F}_+$, or $N_\pi(y)$ is transversal to J at y for a $\tilde{\nu}$-positive collection of $y\in\tilde{F}_+$. We finish the proof of 4.1 by assuming that

$$\tilde{\nu}\left(\left\{y\in\tilde{F}_+\big|\, W_1(y) \text{ is transversal to } J\right\}\right)>0. \tag{4.13}$$

The proof in the other case (when $W_1(y)$ is replaced by $N_\pi(y)$ in (4.13)) is analogous.

We have now reached the point in the proof where the celebrated "Fundamental Theorem for algebraic semi-dispersive billiards", i. e. Theorem 4.4 of [B-Ch-Sz-T(2002)] comes to play. More precisely, we will be using the so called "transversal" version of that theorem to construct a positively measured bundle of local stable manifolds, being transversal to the manifold J. Note that this "transversal" generalization of Theorem 4.4 of [B-Ch-Sz-T(2002)] goes along the same lines as the transversal generalization (i. e. Theorem 3.6 in [K-S-Sz(1990)]) of Lemma 3 of [S-Ch(1987)]. The transversal version of Theorem 4.4 of [B-Ch-Sz-T(2002)] says that there exists a number $\epsilon_1>0$ such that

$$\begin{aligned}&\tilde{\nu}\Big(\big\{y\in\tilde{F}_+\big|\, W_1(y) \text{ is transversal to } J \text{ and}\\ &\gamma_1^{(s)}(y) \text{ has maximum dimension and internal radius } \geq\epsilon_1\big\}\Big)>0.\end{aligned} \tag{4.14}$$

Here $\gamma_1^{(s)}(y)=\gamma_1^{(s)}(y_1)$ denotes the local stable manifold of y_1 in the sub-billiard dynamics $\{S_1^t\}$. The maximum dimension of $\gamma_1^{(s)}(y)$ (mentioned in (4.14)) is $(|P_1|-1)\nu-1$ $(\geq\nu-1)$. Note that, necessarily, we have $|P_1|\geq 2$, otherwise the space $W_1(y)$ of (4.13) could not be transversal to J. Denote the positively measured set featuring (4.14) simply by A. In virtue of the absolute continuity of the foliation $\gamma_1^{(s)}(\,\cdot\,)$, see [K-S(1986)], we have the inequality

$$\mu\left(\bigcup_{y\in A}\gamma_1^{(s)}(y)\right)>0, \tag{4.15}$$

where $\mu=\mu_{\vec{m},r}$ is the Liouville measure of the N-ball system $(\mathbf{M},\{S^t\}_{t\in\mathbb{R}},\mu)$. For the phase points $z\in\bigcup_{y\in A}\gamma_1^{(s)}(y)$ let us construct the modified forward orbit $\{S_*^t z\}_{t\geq 0}$ by deleting all the interactions between particles belonging to different classes of π, thus allowing such particles to overlap. Observe that the phase points z of the set $\bigcup_{y\in A}\gamma_1^{(s)}(y)$ enjoy the following property: If the labels i and j belong to different classes of the partition π, then $\operatorname{dist}\big(q_i^*(t),\, q_j^*(t)\big)\geq 2r-\epsilon_1$ for all $t\geq 0$

on the modified forward orbit $\{S_*^t z\}_{t\geq 0}$. However, according to Theorem 5.1 of [S-Sz(1999)] (or, Main Lemma 4.4 of [Sim(2002)]) this phenomenon cannot occur on a positively measured set of (4.15). Note that, formally, the cited results preclude the positively measured occurence of dist $\left(q_i^*(t),\, q_j^*(t)\right) \geq 2r$, instead of $2r-\epsilon_0$, but this little improvement of the quoted results can be achieved easily by a very small, basically notational, change in their proofs. The obtained contradiction completes the proof of Main Lemma 4.1. □

Finishing the proof of Theorem 2/a. We need to prove (1) of the theorem by using the induction hypothesis. Main Lemma 4.1 asserts that for $\tilde{\nu}$-almost every phase point $x \in \mathcal{SR}^+$ the following properties hold true:

(i) $S^{(0,\infty)}x$ is non-singular, and
(ii) $S^{(0,\infty)}x$ contains infinitely many consecutive, connected collision graphs.

According to Main Lemma 3.1 above (note that this is already the second use of that lemma in the proof of Thorem 2/a), for any $x \in \mathcal{SR}^+$ with a combinatorially rich and non-singular orbit segment $S^{[0,b]}x$, any exceptional, codimension-one manifold NS(Σ) ($\Sigma = \Sigma\left(S^{[0,b]}x\right)$ is the symbolic collision sequence of $S^{[0,b]}x$) cannot even locally coincide with any smooth component of $\mathcal{SR}^+$, thus the full forward orbit $S^{(0,\infty)}x$ of $\tilde{\nu}$-almost every $x \in \mathcal{SR}^+$ is non-singular and sufficient, as claimed. The proof of Theorem 2/a is now complete. □

Concluding Remark. The status of ergodicity in the case $\nu \geq 3$.

What goes wrong with the proof of Theorem 1 in the case $\nu \geq 3$? A key step in proving that theorem was the use of the results of [Sim(2003)], please see 9.3 at the end of that paper. However, the machinery of sections 4–8 of that paper (i. e. proving the non-existence of the so called separating, exceptional manifolds J, see the paragraph in §3 of [Sim(2003)] containing the properties (0)–(3) of J) essentially uses the assumption $\nu = 2$, as this circumstance was pointed out several times thruoghout those sections. Thus, the only thing that hinders the proof of ergodicity in the case $\nu \geq 3$ is a potential survival of the ominous separating, exceptional manifolds J. These are the only instruments capable of separating distinct, open ergodic components, as described in §3 of [Sim(2003)] listing the properties (0)–(3) of J.

References

[B-Ch-Sz-T(2002)] P. Bálint, N. Chernov, D. Szász, I. P. Tóth, *Multidimensional semidispersing billiards: singularities and the fundamental theorem*, Ann. Henri Poincaré **3, No. 3 (2002)**, 451–482.

[B-F-K(1998)] D. Burago, S. Ferleger, A. Kononenko, *Uniform estimates on the number of collisions in semi-dispersing billiards*, Annals of Mathematics **147 (1998)**, 695-708.

[C-H(1996)] N. I. Chernov, C. Haskell, *Nonuniformly hyperbolic K-systems are Bernoulli*, Ergod. Th. & Dynam. Sys. **16 (1996)**, 19-44.

[E(1978)] R. Engelking, *Dimension Theory*, North Holland.

[G(1981)] G. Galperin, *On systems of locally interacting and repelling particles moving in space*, Trudy MMO **43 (1981)**, 142-196.

[K-S(1986)] A. Katok, J.-M. Strelcyn, *Invariant Manifolds, Entropy and Billiards; Smooth Maps with Singularities*, Lecture Notes in Mathematics **1222**, Springer, 1986.

[K-S-Sz(1989)] A. Krámli, N. Simányi, D. Szász, *Ergodic Properties of Semi–Dispersing Billiards I. Two Cylindric Scatterers in the 3–D Torus*, Nonlinearity **2 (1989)**, 311–326.

[K-S-Sz(1990)] A. Krámli, N. Simányi, D. Szász, *A "Transversal" Fundamental Theorem for Semi-Dispersing Billiards*, Commun. Math. Phys. **129 (1990)**, 535–560.

[K-S-Sz(1991)] A. Krámli, N. Simányi, D. Szász, *The K–Property of Three Billiard Balls*, Annals of Mathematics **133 (1991)**, 37–72.

[K-S-Sz(1992)] A. Krámli, N. Simányi, D. Szász, *The K–Property of Four Billiard Balls*, Commun. Math. Phys. **144 (1992)**, 107-148.

[O-W(1998)] D. Ornstein, B. Weiss, *On the Bernoulli Nature of Systems with Some Hyperbolic Structure*, Ergod. Th. & Dynam. Sys. **18 (1998)**, 441-456.

[Sim(1992)-I] N. Simányi, *The K-property of N billiard balls I*, Invent. Math. **108 (1992)**, 521-548.

[Sim(1992)-II] N. Simányi, *The K-property of N billiard balls II*, Invent. Math. **110 (1992)**, 151-172.

[Sim(2002)] N. Simányi, *The Complete Hyperbolicity of Cylindric Billiards*, Ergodic Th. & Dyn. Sys. **22 (2002)**, 281-302.

[Sim(2003)] N. Simányi, *Proof of the Boltzmann-Sinai Ergodic Hypothesis for Typical Hard Disk Systems*, Inventiones Mathematicae **154, No. 1 (2003)**, 123-178.

[Sim(2004)-I] N. Simányi, *Proof of the Ergodic Hypothesis for Typical Hard Ball Systems*, Annales Henri Poincaré **5 (2004)**, 203-233.

[Sim(2004)-II] N. Simányi, *Proving the Ergodic Hypothesis for Billiards With Disjoint Cylindric Scatterers*, Nonlinearity **17 (2004)**, 1-21.

[Sin(1963)] Ya. G. Sinai, *On the Foundation of the Ergodic Hypothesis for a Dynamical System of Statistical Mechanics*, Soviet Math. Dokl. **4 (1963)**, 1818-1822.

[Sin(1970)] Ya. G. Sinai, *Dynamical Systems with Elastic Reflections*, Russian Math. Surveys **25:2 (1970)**, 137-189.

[S-Ch(1987)] Ya. G. Sinai, N.I. Chernov, *Ergodic properties of certain systems of 2–D discs and 3–D balls*, Russian Math. Surveys **42, No. 3 (1987)**, 181-207.

[S-Sz(1999)] N. Simányi, D. Szász, *Hard ball systems are completely hyperbolic*, Annals of Mathematics **149 (1999)**, 35-96.

[S-Sz(2000)] N. Simányi, D. Szász, *Non-integrability of Cylindric Billiards and Transitive Lie Group Actions*, Ergod. Th. & Dynam. Sys. **20 (2000)**, 593-610.

[Sz(1994)] D. Szász, *The K-property of 'Orthogonal' Cylindric Billiards*, Commun. Math. Phys. **160 (1994)**, 581-597.

[V(1979)] L. N. Vaserstein, *On Systems of Particles with Finite Range and/or Repulsive Interactions*, Commun. Math. Phys. **69 (1979)**, 31-56.