Chaotic and Ergodic Properties of Cylindric Billiards

Péter Bálint
Alfréd Rényi Institute of Mathematics,
Hungarian Academy of Sciences
H-1364, Budapest, P.O.B. 127, Hungary
E-mail: bp@math-inst.hu

June 11, 1999

Abstract

Chaotic and ergodic properties are discussed in this paper for various subclasses of cylindric billiards. Common feature of the studied systems is that they satisfy a natural necessary condition for ergodicity and hyperbolicity, the so called transitivity condition. Relation of our discussion to former results on hard ball systems is twofold. On the one hand, by slight adaptation of the proofs we may discuss hyperbolic and ergodic properties of 3 or 4 particles with (possibly restricted) hard ball interactions in any dimensions. On the other hand a key tool in our investigations is a kind of connected path formula for cylindric billiards, which, together with the conservation of momenta, gives back, when applied to the special case of Hard Ball Systems, the classical Connected Path Formula.

Keywords: Cylindric Billiards, Hard Ball Systems, Ergodicity, Hyperbolicity.

1 Introduction

One of the most interesting open questions of statistical physics and dynamical systems theory is the so-called Boltzmann-Sinai Ergodic hypothesis, i.e. the conjecture that Hard Ball Systems in physical dimensions are ergodic (as to the history of this conjecture see [22]). In addition to its physical relevance the question is interesting as it is highly non-trivial from the mathematical point of view. Nevertheless, there is an increasing belief in mathematical physics communities that the proof is within reach. Thus a natural question is the following: what could be a suitable category of dynamical systems containing all hard ball systems for which transparent necessary and sufficient conditions for ergodicity can be proven?

Dynamical properties of hard ball systems are mainly characterized by the fact that they belong to the category of semi-dispersive billiards. Nevertheless, semi-dispersive billiards in their full generality may show a too extreme variety of dynamical behaviour to guess the conditions for ergodicity. It has turned out that one possible choice for the systems to answer the question above is the class of *cylindric billiards.* (This class of dynamical systems was introduced in [20]; as to conjectures related to its ergodic-hyperbolic properties see [22], [19] or conjecture 2.1 in the present paper.)

The purpose of our study is to prove ergodicity for cylindric billiards with a low number of scatterers (or, more precisely, with cylinders at most a low number of which may have generator subspaces with nontrivial intersection, see Theorem 2.4). For some systems (Theorem 2.5) instead of ergodicity only the weaker hyperbolic/chaotic property is proven (from which by [6] it follows that there are at most a countable number of ergodic components, on each of which the dynamics possesses the K- and the B-properties). The most difficult steps of the proofs (from the geometric-algebraic considerations) are discussed by the help of a method for the calculation of neutral subspaces (Lemma 2.7) which can be viewed as an analogue of the Connected Path Formula for Hard Ball Systems ([13]) in this cylindric billiard setting.

The paper is organized as follows. In Section 2 we summarize the most important prerequisites, state the results and make some general remarks on the proofs. The proof of the above mentioned Lemma 2.7 is also presented here (subsection 2.3). In Section 3 ergodicity and hyperbolicity is proven for cylindric billiards with pairwise transversal generator subspaces (Theorem 2.4). In Section 4 we show hyperbolicity for billiards with three cylinders (Theorem 2.5). The sections consist of further subsections according to the steps of the proofs. In the Appendix we summarize how our results are applicable to some particle systems with hard ball interactions (together with some further simple generalizations).

2 Prerequisites and general observations

2.1 Definition of the dynamical system and summary of results

The subject of our study, the category of cylindric billiards is a simple subclass of semi-dispersive billiards – as to the notions in connection with these systems see [8]. In our case the configuration space of the billiard is defined by cutting out a finite number of cylindric regions from the d-dimensional unit torus, i.e. $Q = \mathbf{T^d} \setminus (C_1 \cup \cdots \cup C_k)$. For the precise definition of the cylinders we need three data for each C_i. We fix A_i, a subspace of the d-dimensional Euclidean space $\mathbf{R^d}$, the so-called *generator subspace* of the cylinder. A_i should be a so-called lattice-subspace to get a properly defined cylinder on $\mathbf{T^d}$ after factorization ([19]). We assume $dim(L_i) \geq 2$, where $L_i = A_i^\perp$ is the notation for the *base subspace*, the orthogonal complement of A_i. The *base*, $B_i \subset L_i$ is a convex, compact domain, for which, to ensure semi-dispersivity, the C^2-smooth boundary ∂B_i is assumed to have everywhere positive definite second fundamental form. Furthermore a translational vector $t_i \in \mathbf{R^d}$ is given to place our cylinder in $\mathbf{T^d}$. By the help of these data our cylinders are defined as:

$$C_i := \{a + l + t_i : a \in A_i, l \in B_i\}/\mathbf{Z^d}. \tag{2.1}$$

To avoid possible complications we assume that: (i) the domain B_i does not contain any pair of points congruent modulo $\mathbf{Z^d}$; (ii) the interior of the configuration space $Q = \mathbf{T^d} \setminus (C_1 \cup \cdots \cup C_k)$ is connected.

It is time to give the definition of our dynamical system $(M, S^{\mathbf{R}}, \mu)$. Our $2d-1$-dimensional phase space is the unit tangent bundle of Q, i.e. $M = Q \times \mathbf{S^{d-1}}$ (here $\mathbf{S^{d-1}}$ is the $d-1$-dimensional unit sphere). The dynamics $S^t x$ for a phase point $x \in M$ is understood in continuous time and defined by uniform motion inside the domain and specular reflections at the boundary (the cylinders). Finally, μ is, as usual, the Liouville-measure (i.e. $d\mu = const\, dq dv$), which is invariant for the flow. For future convenience we fix here some more notation. A finite trajectory segment, $S^{[a,b]}x$, is the collection of points $S^t x$ on the trajectory of x for which $a \le t \le b$. For any phase point $x = (q, v) \in Q \times \mathbf{S^{d-1}} = M$ the natural projections are defined as $p(x) = v$ and $\pi(x) = q$. By a non-trivial sub-billiard of our cylindric billiard we mean the billiard dynamical system we get by cutting out only some of the cylinders C_i.

Example. Consider N ball particles with finite masses and radii moving on the ν-dimensional torus. To define dynamics, assume furthermore that a so called *collision graph* (or graph of interactions), $\Gamma = (\mathcal{V}, \mathcal{E})$ is given. Here $\mathcal{V}$ is the finite set of the N particles. Pairs contained among the edges $\mathcal{E}$ do interact via hard ball collisions, while non-connected pairs do not interact at all (see also [17]); otherwise the dynamics is governed by free motion. In the case of a complete collision graph we get the so deeply studied Hard Ball Systems. This dynamical system (restricted to the constant value submanifold of the trivial integrals of motion) is equivalent to a cylindric billiard with configuration space in $\mathbf{T^{N\nu-\nu}}$. The cylinders we cut out correspond to the allowed interactions (in case the configurational domain Q is not connected – i.e. the radii of the balls are not small enough – we may view the finitely many connected components of Q as configurational domains for independent billiard systems). To get a detailed description of this isomorphism see [18] and references therein. Thus cylindric billiards are indeed generalizations of all possible Hamiltonian systems defined on tori with restricted hard ball interactions. As to the relevance of our results to this example see the Appendix.

Important Remark. If we want to describe Hard Ball Systems as cylindric billiards we have to be a bit more precise. Instead of $\mathbf{R^d}/\mathbf{Z^d}$ we should allow for the more general $\mathbf{T^d} = \mathbf{R^d}/\mathcal{L}$ at the definition of the configurational domain. (Here $\mathcal{L}$ is a lattice in the Euclidean plane $\mathbf{R^d}$). Throughout the definition of the dynamical system $\mathbf{Z^d}$ should be replaced by the lattice $\mathcal{L}$ (e.g. formula (2.1)).
The results of this paper are formulated and proved for cylindric billiards on $\mathbf{R^d}/\mathbf{Z^d}$. Nevertheless, it is important to note that for a large class of general lattices $\mathcal{L}$ the proofs go through without modification. More precisely, the dynamical-topological considerations (subsections 3.2 and 4.2) work for exactly those lattices $\mathcal{L}$ that have the following

property: *for any subspace A which is a lattice subspace with respect to the lattice $\mathcal{L}$, the orthogonal complement $L = A^\perp$ should be a lattice subspace as well.* (for a detailed analysis and equivalent formulations of this property see [14]).
As to particle systems with hard ball interactions the corresponding cylindric billiard is defined with a lattice of the above type whenever *all mass ratios of the ball particles are rational.* Thus the results of the paper are directly applicable to (the suitable subclasses of) Hard Ball Systems whenever the mass ratios are rational. However, in case of a cylindric billiard equivalent to a system with hard ball interactions, it is possible to use an argument at the dynamical-topological considerations (see [18, 13]) which is much different from the proofs of Propositions 3.3 and 4.2 in this paper and which does not rely on the properties of $\mathcal{L}$. Thus the present results work even if there are irrational mass ratios, as the rest of the proof (subsections 3.1 and 4.1) does not use any property of $\mathcal{L}$. (A much more detailed study of these and related issues is to be found in [14]).

The two most important phenomena characterizing the dynamics in semi-dispersive billiards is on the one hand that they enjoy some *hyperbolicity* (in fact, there exists an invariant, but not necessarily strictly invariant cone field, cf. [11]), and on the other hand that *singularities* are present. As to the former one, among the most interesting questions of the theory is whether hyperbolicity is strong enough to ensure that our dynamical system is (i) completely **hyperbolic**, i.e. with respect to the invariant measure μ almost everywhere all relevant Lyapunov-exponents of the flow are nonzero (such dynamical systems are sometimes referred to as chaotic); (ii) **ergodic** with respect to the measure μ.
There are two possible types of singularities in billiards. A collision at the boundary point $(q, v) \in \partial M$ is said to be *multiple* if at least two smooth pieces of the boundary ∂Q meet at q, and is *tangential* if the velocity v is tangential to ∂Q at q. At tangential reflection points the dynamics is continuous, though not smooth. However, at a multiple reflection point the dynamics is not even continuous. Thus the future semi-trajectory (or the outgoing velocity) is not well-defined for a multiple reflection point – for such points two *trajectory branches* can be considered as the limits of the smooth dynamics. We shall denote the set of all singular reflection points (belonging to any of the above two types, in case of multiple collision supplied with outgoing velocity v^+) by $\mathcal{SR}^+$.
Before introducing some further notation it is time to mention some key prerequisites. As it is described in [1], it is an important feature of generic semi-dispersive billiards that there are no finite trajectory segments with an infinite number of collisions on them (analogous statements have first appeared in [24, 4]). On the other hand, it is possible that some trivial subsets are present for the points of which the trajectories do not collide at all. Such points lie on a finite number of one codimensional submanifolds (for details see subsection 3.2 in this paper and, especially, the Appendix of [21]). The complementary set, $M^\#$, contains only such typical trajectories for which infinitely many collisions are present – the following characterization is related to this set. $M^* \subset M^\#$ is the set of phase points whose trajectories contain infinitely many collisions such that at most one is singular among them. $M^0 \subset M^*$ is the set of regular phase points — i.e.

whose entire trajectory (with infinitely many collisions on it) avoids $\mathcal{SR}^+$; for the points in $M^1 = M^* \setminus M^0$ there is exactly one singular reflection. To measure the size of these sets we use the small inductive topological dimension (for details see [23, 5]). It is not hard to see that $\mathcal{SR}^+$ is a codimension 1 subset of ∂M, and, as a consequence, M^0 is of full measure in M, while $M^\# \setminus M^*$ is of codimension 2 (cf. [8]).

The main philosophy of the study of semi-dispersive billiards has always been the principle that in some sense *hyperbolic behaviour is stronger than the effect of singularities.* According to this philosophy we expect that dynamics in a cylindric billiard is chaotic and ergodic unless some geometric degeneracy of the cylinders is present. This expectation is formulated precisely in the conjecture below, for which we need some more preparation.

For any cylinder C_i let $\mathcal{G}_i$ be the group of all orientation preserving orthogonal transformations in $\mathbf{R^d}$ that leave the points of the generator subspace A_i fixed. Denote furthermore by $\mathcal{G}$ the group of transformations algebraically generated by all such groups $\mathcal{G}_i : i = 1, ..., k$. Observe that the group $\mathcal{G}$, being a subgroup of the special orthogonal group $SO(d)$, has a natural action on $\mathbf{S^{d-1}}$ (and thus on $\mathbf{R^d}$ as well).

Conjecture 2.1 *The cylindric billiard is completely hyperbolic and moreover ergodic if and only if the group action of $\mathcal{G}$ is transitive on the unit velocity sphere $\mathbf{S^{d-1}}$. From here on we refer to the transitivity of the action of $\mathcal{G}$ as to the transitivity condition on our cylinders.*

The remarks below are discussed in detail in [19].

Remark 2.2 *For the case of restricted hard ball interactions (see the example above) the transitivity condition for the isomorphic cylindric billiard is satisfied if and only if the collision graph is connected. Thus if one were able to prove the above conjecture one would immediately get ergodicity and hyperbolicity for a large class of particle systems with hard ball interactions, including the so much discussed Hard Ball System case.*

Remark 2.3 *The equivalence of the following three conditions on the cylindric billiard is demonstrated in [19]:*
(i) The action of $\mathcal{G}$ is transitive on $\mathbf{S^{d-1}}$
(ii) The action of $\mathcal{G}$ is irreducible on $\mathbf{R^d}$
(iii) The system of base subspaces $L_1, \cdots, L_k$ has the Orthogonal Non-Splitting Property, i.e. there is no orthogonal splitting $\mathbf{R^d} = K_1 \oplus K_2$ for which $dim(K_j) > 0$ and which has the property that for any $i = 1, \cdots, k$ either $L_i \subset K_1$ or $L_i \subset K_2$.

As a consequence of remark 2.3 we immediately see that *the transitivity condition is necessary* for both ergodicity and hyperbolicity. Indeed, let us suppose that the transitivity condition does not hold. This in turn implies by virtue of (iii) above that there exists a nontrivial orthogonal splitting $\mathbf{R^d} = K_1 \oplus K_2$. Then it is easy to see that for any phase point $(q, v) \in M$ the two quantities $\|P_{K_j}(v)\|$ remain constant under the time evolution, thus nontrivial integrals of motion are present (here and from here on $P_K(v)$ stands for the orthogonal projection of the vector v onto the subspace K, while $\|z\|$ is the absolute value of the vector z).

In contrast to the necessity of the condition, proving that the transitivity of the action is indeed sufficient for both ergodicity and hyperbolicity is an extremely nontrivial task. (By remark 2.2, proving conjecture 2.1 in its full generality would imply the ergodicity and the hyperbolicity of Hard Ball Systems, a problem that has been a subject of active research for the last couple of decades.) The results in this paper are proofs of the conjecture for some subclasses of cylindric billiards. In the rest of this subsection we give a summary of results.

The simplest possible class of cylindric billiards one could imagine would be the case $A_i = \{0\}$ for all the cylinders C_i. For such systems the scatterers are strictly convex, the billiard is dispersive, thus ergodicity and hyperbolicity follow as consequences of the local hyperbolic and ergodic theorems (see [8] or [2]). On the other hand the transitivity condition is naturally satisfied.
To get more complicated cylindric billiards we must allow for 'thicker' generator subspaces. The more cylinders may have generator subspaces with nontrivial intersection, the more difficult it is to handle the system. Theorem 2.4 below proves the desired properties for the simplest possible nontrivial class, while Theorem 2.5 gives the hyperbolicity of some more 'complicated' systems. In a sense the complexity of the increasing number of generator subspaces with nontrivial intersection corresponds to the increasing number of particles in Hard Ball Systems. Indeed, the system of two balls on the torus is equivalent to a dispersive billiard (cf. [8]); while, as it is discussed in the remarks below, the conditions of Theorems 2.4 and 2.5 are valid for Hamiltonian systems of 3 or 4 ball particles, respectively.

Theorem 2.4 *Let us consider a cylindric billiard with an arbitrary number of scatterers* $C_1, \cdots, C_k$
(i) which satisfy the transitivity condition;
(ii) for which it is true that for any two cylinders C_i, C_j $(i \neq j)$ *the corresponding generator subspaces are transversal:* $A_i \cap A_j = \{0\}$.
The dynamical system is ergodic and hyperbolic. (Moreover, the dynamics is K-mixing ([6]) and possesses the Bernoulli property ([3, 12])).

Remarks. This result is a natural generalization of the one discussed in [16]: the ergodicity and hyperbolicity of a billiard with two cylinders was demonstrated there, with one additional assumption on the scatterers besides our (i)-(ii).
The result of Theorem 2.4 implies the ergodicity of three particles, both for the classical case of three hard balls with positive radii and for the case when two of the three particles interact only with the third particle and not with each other as they have zero radius.

Theorem 2.5 *Let us consider a billiard with three cylinders which satisfy the transitivity condition.*
The dynamical system is hyperbolic.

Remarks. As for a particle system that belongs to this class one may consider four particles only one of which has nonzero radius (in this model the particles with zero radius do not interact; we have three pair interactions, i.e. three cylinders).
It is a natural question what one can do if all the four particles have nonzero radius. A slight adaptation of our methods gives hyperbolicity for this case as well (as it is demonstrated in the Appendix). Although for four hard balls even ergodicity has been proven ([10]), the new results are interesting as they are valid without dimensional restriction, i.e. for two dimensional disks as well.

One more remark is in order. It may happen that two cylinders are parallel, i.e. for a pair $i \neq j$ $A_i = A_j$. In such a case collisions with C_i and C_j have exactly the same effect on the dynamics, thus the two cylinders can be considered as identical. We have formulated the theorems above and will go on with the proofs below by assuming that such a parallelity does not occur. Nevertheless the results remain trivially true if we state the conditions on such identified classes of parallel cylinders rather than on the cylinders themselves.

2.2 Basic strategy of the proofs

In this subsection some 'traditional' concepts are summarized that play an essential role in the theory of semi-dispersive billiards. Our discussion is very brief, for more details see the literature, especially [19, 8, 10]. Let us consider a nonsingular finite trajectory segment $S^{[a,b]}x$, where $a < 0 < b$ and a, b, 0 are not moments of collision.
$\mathcal{N}_0(S^{[a,b]}x)$, the **neutral subspace** at time 0 for the segment $S^{[a,b]}x$ is defined as follows:

$$\begin{aligned}\mathcal{N}_0(S^{[a,b]}x) := \{ \quad & w \in \mathbf{R^d} : \exists(\delta > 0) s.t. \forall \alpha \in (-\delta, \delta) \\ & p(S^a(q(x) + \alpha w, v(x))) = p(S^a x) \& \\ & p(S^b(q(x) + \alpha w, v(x))) = p(S^b x)\}.\end{aligned}$$

Observe that $v(x) \in \mathcal{N}_0(S^{[a,b]}x)$ is always true, the neutral subspace is at least 1 dimensional. Neutral subspaces at time moments different from 0 are defined by $\mathcal{N}_t(S^{[a,b]}x) := \mathcal{N}_0(S^{[a-t,b-t]}(S^t x))$, thus they are naturally isomorphic to the one at 0.

The following notion is one of the most important concepts in the theory of semi-dispersive billiards. The non-singular trajectory segment $S^{[a,b]}x$ is **sufficient** if for some (and in that case for any) $t \in [a, b]: \quad dim(\mathcal{N}_t(S^{[a,b]}x)) = 1$. A point $x \in M^0$ is said to be sufficient if its entire trajectory $S^{(-\infty,\infty)}x$ contains a finite sufficient segment. Singular points are treated by the help of trajectory branches (see [8]): a point $x \in M^1$ (this precisely means that the entire trajectory contains one singular reflection) is sufficient if both of its trajectory branches are sufficient.

Sufficiency has a picturesque meaning; roughly speaking a trajectory segment is sufficient if it has encountered all degrees of freedom. Nevertheless the concept is important as very strong theorems hold in open neighborhoods of sufficient points (on more general formulations of sufficiency and on the local ergodicity theorem see [11, 2, 8]).
Local Hyperbolicity Theorem. Every sufficient phase point $x \in M$ has an open

neighborhood $x \in U \subset M$, such that for μ a.e. $y \in U$ the relevant Lyapunov exponents of the flow are nonzero.

Local Ergodicity Theorem or Fundamental Theorem of Semi-Dispersive Billiards. Assume that some geometric conditions are true for the singularities of our semi-dispersive billiard (most importantly the *Chernov-Sinai Ansatz* holds, see e.g. [8]). Then every sufficient phase point $x \in M$ has an open neighborhood $x \in U \subset M$ which belongs to one ergodic component.

Now let us examine the trajectory segment $S^{[a,b]}x$ from another point of view. We denote with τ_j; $a < \tau_1 < \cdots < \tau_n < b$ the moments of collision on the segment, i.e. for any $j = 1, ..., n$ it is true that $S^{\tau_j}x = (q_j, v_j) \in \partial M$. Let us assume $q_j \in \partial(C_{l(j)})$, i.e. at time τ_j reflection occurs at the cylinder $C_{l(j)}$. The finite sequence of symbols $(l(1), \cdots, l(n))$ is called the *symbolic collision sequence* of the trajectory segment.

A finite trajectory segment is said to be *connected* if it holds for the corresponding symbolic collision sequence that the system of subspaces $(A_{l(1)}, \cdots, A_{l(n)})$ is transitive in the sense as it is formulated at conjecture 2.1.

A finite trajectory segment is said to be *rich* if the corresponding symbolic collision sequence contains 'enough collisions' in some combinatorical sense. Formulating the concept of richness so vaguely is done by purpose: the useful notion of richness depends and has always depended on the specific model one considers. Furthermore, the useful formulation also depends on what one wants to prove; for proving ergodicity definitely a stronger notion of richness is needed than for showing hyperbolicity. Nevertheless, the concepts of connectedness and richness should be related in a way. With any notion of richness being used we call a phase point rich/connected if it has a rich/connected finite trajectory segment and we call it poor if it is not rich.

It is time to make a general remark on a type of trivially sufficient (and, simultaneously, rich) sequences.

Remark 2.6 *Let us assume that among the cylinders there exists a C_i for which the generator space is trivial, i.e. $A_i = \{0\}$. It is easy to see that any finite trajectory segment for which a collision with C_i occurs is sufficient. Such segments will always be considered to be rich. In our specific examples we shall not take care of these trivially rich sequences as they always have very good hyperbolic properties.*

Dynamics of a phase point in a finite time interval is mainly characterized by its symbolic collision sequence. It may happen, however, that trajectories with different collision sequences show the same dynamical behaviour. For example two consecutive collisions with the same cylinder have the same effect as if only one collision occurred. An *island* is a maximal subsequence of a collision sequence consisting of consecutive collisions with the same cylinder (see [16, 9]). Collision sequences with the same island structure may be treated as equivalent. We shall turn back to the question of equivalence and to the correct formulation of richness at the end of subsection 2.3.

Following tradition (e.g. [16, 10]) proving either hyperbolicity or ergodicity is done in several steps. For the proof of hyperbolicity one needs

(H1) *Geometric-algebraic considerations* to prove that μ-a.e. rich point in the phase space is sufficient; and
(H2) *Dynamical-topological considerations* to prove that μ-a.e. point of the phase space is rich.
We get sufficiency μ-almost everywhere if (H1) and (H2) hold simultaneously, this together with the above mentioned local hyperbolicity theorem implies the hyperbolicity of the dynamics.

Ergodicity is a more difficult task as one has to ensure that the local ergodic components indeed do make up one ergodic component (i.e. they are not separated by codimension 1 subsets). It has to be shown that the 'exceptional sets' are, besides being of zero μ-measure, slim subsets of the phase space (a subset is slim if it can be covered by a countable union of closed zero measure sets of topological codimension 2, for a detailed analysis see the recent review [23]). A further difficulty is that one has to consider the singular orbits (they can be ignored in the proof of hyperbolicity as they form a zero measure set (a set of codimension 1)). The steps of the proof of ergodicity are:
(E1) *Geometric-algebraic considerations* to prove that the set of points that are rich and non-sufficient is slim;
(E2) *Dynamical-topological considerations* to prove that the set of poor (and, simultaneously, non-sufficient) points is slim; and
(E3) Considerations on the *singularities* of the system, including the verification of the Chernov-Sinai Ansatz.
Putting parts (E1-E3) together implies, on the basis of the local ergodicity theorem, the ergodicity of the billiard dynamics.

2.3 Calculation of neutral subspaces

In the study of various cylindrical billiards (especially hard ball systems, see e.g. [10, 18]) geometric-algebraic considerations are usually much more difficult than the rest of the proof. This is true for our results as well – as the reader shall see from sections 3 and 4, while dynamical-topological considerations are more or less easy adaptations of methods discussed in the literature, some new ideas are needed in the geometric-algebraic part.

Our task in the geometric-algebraic part is approximately the following. Let us assume a trajectory segment with one particular rich symbolic collision sequence is given (there is always a finite number of symbolic collision sequences one has to study, depending on the specific notion of richness we use for the model). If we knew how the neutral subspace at a given (non-collision) time moment of the segment explicitly looks like, we could conclude that the neutral subspace is one dimensional (i.e. sufficiency holds) unless some degeneracy of the projections of velocity vectors at different time moments occurs. This way non-sufficient phase points with a given rich collision sequence can be described by some algebraic equations that characterize the above mentioned degeneracy. These sets are, in general, algebraic varieties. For brevity, however, throughout the paper we shall abuse the terminology and refer to them as submanifolds. In order to describe these exceptional sets it is of key importance to find explicit methods for the calculation of neutral subspaces of different collision sequences.

As to Hard Ball Systems, remarkable progress has been achieved related to this problem through the so called *Connected Path Formula* invented by N. Simányi (see [13]). This formula however, which has been so powerfully used since its invention (see e.g. [18, 13]), uses the properties of hard ball systems (the conservation of momenta), thus it cannot be applied directly to the wider class of general cylindric billiards. In this subsection some general observations on the calculation of neutral subspaces are made which will be of further use in the rest of the paper playing the role of the connected path formula. Actually, when applied to the isomorphic cylindric billiard in the case of particle systems with hard ball interactions, our Lemma 2.7, together with the conservation of momenta, gives back the classical Connected Path Formula.

From here on $S^{[a,b]}x$ is a fixed trajectory segment with collision sequence $(l(1), ...l(n))$; the moments of collision we denote by $a < \tau_1(x) < ... < \tau_n(x) < b$, while $t \in [a, b]$ is an arbitrary non-collision time moment. $v(t)$ stands for the velocity vector at time t and $w(t)$ for an arbitrary fixed vector of the neutral subspace, also calculated at time t (we use the natural isomorphism of neutral subspaces at different time moments mentioned in subsection 2.2). As before, $P_K(w)$ denotes the orthogonal projection of the vector w onto the subspace K.
For all the collisions $i = 1, ..., n$ we define linear functionals $\alpha_i : \mathcal{N}_t(S^{[a,b]}x) \mapsto \mathbf{R}$, the *advances* of the collisions (see e.g. [13]) as the unique linear extensions of the linear functions α'_i defined in an open neighborhood of the origin of $\mathcal{N}_t(S^{[a,b]}x)$:

$$\alpha'_i(w(t)) := \tau_i(x) - \tau_i(S^{-t}T_{w(t)}S^t x).$$

Here $T_{w(t)}(x) = T_{w(t)}(q, v) = (q + w(t), v)$ denotes the pure spatial translation by the vector $w(t)$. By the natural isomorphism of neutral subspaces the value of an advance does not depend on the time moment when the neutral vector is calculated.

It is also easy to conclude (see e.g. [16]) that for any t such that $\tau_{i-1} < t < \tau_i$ or $\tau_i < t < \tau_{i+1}$:

$$w(t) = \alpha_i P_{L_{l(i)}}(v(t)) + a \tag{2.2}$$

where $a \in A_{l(i)}$ is an element of the generator subspace which (just like the value of the advance for the neutral vector w, α_i) does not depend on time.

Lemma 2.7 *We use the notations introduced above, thus our collision sequence is* $(l(1), ..., l(n))$, *collision moments are denoted by* τ_i, *the advances by* α_i, $i = 1, ..., n$ *(more precisely, the* α_i*-s are the values of the advances for a fixed neutral vector). Non-collision time moments* t_i *are fixed between the consecutive collisions,* $\tau_i < t_i < \tau_{i+1}$. *Our aim is to calculate the neutral subspace at time* t_{n-1}, *i.e directly before the last collision. The following equations are valid for any* $j = 1, ..., (n-1)$ *and any neutral vector* $w(t_{n-1})$:

$$\begin{aligned} P_{L_{l(j)}}(w(t_{n-1})) = \quad & (\alpha_j - \alpha_{j+1}) P_{L_{l(j)}}(v(t_j)) + \ ... \ + \\ & (\alpha_{n-2} - \alpha_{n-1}) P_{L_{l(j)}}(v(t_{n-2})) + \alpha_{n-1} P_{L_{l(j)}}(v(t_{n-1})). \end{aligned} \tag{2.3}$$

N.B. For the projection of any such neutral vector onto the base subspace of the last cylinder in the collision sequence by (2.2) it trivially holds:

$$P_{L_{l(n)}}(w(t_{n-1})) = \alpha_n P_{L_{l(n)}}(v(t_{n-1})).$$

Proof. For brevity we introduce the notation: P_i for $P_{L_{l(i)}}$, v^i for $v(t_i)$ and w^i for $w(t_i)$. From (2.2) the following series of equations are straightforward:

$$w^i - w^{i-1} = P_i(w^i) - P_i(w^{i-1}) = \alpha_i P_i(v^i) - \alpha_i P_i(v^{i-1}).$$

We only need one more well-known observation for the proof; the dynamics of a velocity vector is determined by specular reflections, thus while colliding with C_i it's projection onto the generator subspace A_i does not change:

$$P_i(v^i) - P_i(v^{i-1}) = v^i - v^{i-1}.$$

Putting the above two equations together and projecting orthogonally onto *any* L_j, $j = 1, ..., n$ we get:

$$P_j(w^i) = P_j(w^{i-1}) + \alpha_i(P_j(v^i) - P_j(v^{i-1})). \tag{2.4}$$

On the other hand from (2.2) it trivially follows that

$$P_j(w^j) = \alpha_j P_j(v^j).$$

This together with (2.4) summed over $i = (j+1)...(n-1)$ gives the telescopic expression (2.3). $\square$

The advantage of the Lemma is that in case our collision sequence is long enough (i.e. $\Sigma_{i=1}^{n-1} L_{l(i)} = \mathbf{R^d}$), any neutral vector is determined by the values of the advances through the simple equations (2.3).

The already mentioned equivalence of collision sequences is one more issue to be discussed briefly before closing this subsection (an analogous discussion has already appeared in [10], principle 4.1) . We say that two collision sequences are *equivalent* if they can be transformed into each other by
(i) time direction change for the whole sequence, i.e. $(l(1), ..., l(n)) \sim (l(n), ..., l(1))$;
(ii) doubling a collision with the same cylinder, i.e. $(.., l, l, ..) \sim (.., l, ..)$;
(iii) by the interchange of two consecutive collisions if the corresponding cylinders have orthogonal base subspaces; i.e. $(.., l, m, ..) \sim (.., m, l, ..)$ whenever $L_m \perp L_l$ holds.
As it will be clear from the considerations in subsections 3.1 and 4.1, these collision sequences are indeed equivalent from the viewpoint of geometric-algebraic considerations. The fact that the structure of neutral subspaces at the corresponding time moments (i.e. in the cutting time moments, see below) is the same is a straightforward consequence of the definitions and the equations above. The advantage of this notion is that it is enough to consider only the shortest equivalent forms of all possible rich collision sequences.

In the dynamical-topological part, however, one has to be more careful with the equivalence of collision sequences. The following notion, suggested by N. Simányi ([15]), helped a lot in a clearer formulation of the dynamical-topological considerations. In addition to islands we introduce the concept of *archipelago*. An archipelago is a maximal subsequence $(l_1, ..., l_n)$ of a symbolic collision sequence for which it is true that any two cylinders present are either identical ($L_{l(i)} = L_{l(j)}$) or orthogonal ($L_{l(i)} \perp L_{l(j)}$). One more important notion is that of a *cutting time moment* which we define as a non-collision time moment between two consecutive archipelagos. It is an important question whether it is true that any collision sequence has a unique archipelago structure. (By archipelago structure we mean the series of consecutive archipelagos with the cutting time moments in between). As it will not be difficult to see, for the models discussed in sections 3 and 4, this uniqueness indeed holds, which is a specific feature of these cylindric billiards. Given this uniqueness it is straightforward that (modulo time direction change) two collision sequences are equivalent if and only if they have the same archipelago structure.
Now we may turn back to the question of a suitable formulation of richness, which should be invariant under equivalence transformations. (It is straightforward to see that the concept of connectedness depends solely on the archipelago structure). For proving hyperbolicity double connectedness is a possibly good notion of richness in general; i.e. the trajectory segment is rich if there is a cutting time moment t, $a < t < b$, such that both segments $S^{[a,t]}x$ and $S^{[t,b]}x$ are connected (note, however Remark 2.6). In fact, this is the notion of richness we use in section 4.

3 Theorem 2.4 – pairwise transversal generator subspaces

The subject of this section is a cylindric billiard with scatterers $C_1, ..., C_k$ which, as it is formulated in Theorem 2.4, in addition to transitivity condition have the property $A_i \cap A_j = \{0\}$ whenever $i \neq j$. For such a system, given any pair of orthogonal cylinders $L_i \perp L_j$, one surely has: $A_i = L_j$ and $L_i = A_j$. As a cosequence the archipelago structure is unique for any collision sequence. There are two possible types of archipelagos: an archipelago either contains one single island, or it is a maximal sequence of alternating islands of two orthogonal cylinders.
For symbolic collision sequences we use the notation $\Sigma = (a, b, ..)$, here the symbol a denotes the cylinder C_a with generator subspace A_a. Lowercase Latin letters $a, b, c, ..$ denote the cylinders and the corresponding Greek letters $\alpha, \beta, \gamma..$ denote the advances of the collisions, respectively. For vectors w of the neutral subspace and the velocity vector v upper indices refer to the time moments, while lower indices indicate orthogonal projections onto base subspaces. For instance v_a^+ is the orthogonal projection of the velocity vector at time t_+ onto the subspace L_a (nevertheless the symbol $P_{L_a}(v(t_+))$ already introduced in the previous section may be used as well for the same quantity). The notation $K_1 + K_2$ is used for the subspace of $\mathbf{R^d}$ spanned by to given (not necessarily

orthogonal) subspaces K_1 and K_2.

3.1 Geometric-Algebraic Considerations

Before going into details an easy but important observation can be made. Under the conditions of Theorem 2.4 a trajectory segment is sufficient iff there are two consecutive collisions with different cylinders (i.e. two consecutive islands) for which the advances are equal. Indeed, let the two consecutive collisions be (a, b) and let us calculate the neutral vector w in a time moment t between the two collisions, where the velocity is v. By (2.2):

$$w_a = \alpha v_a; \qquad w_b = \alpha v_b$$

where α is the common advance. Now as $A_a \cap A_b = \{0\}$, the above equations completely characterize the vector w. Thus $w = \alpha v$, which is exactly sufficiency.

Proposition 3.1 *Assume $\Sigma = (a, b, c)$ is an arbitrary collision sequence of three collisions that has no shorter equivalent form (the possibility $a = c$ is not excluded). For any regular phase point $x \in M^0$ a bounded trajectory segment of which has collision sequence Σ there exist an open neighborhood $x \in U_0$ and a submanifold N of M such that (i) $codim(N) \geq 1$ and (ii) $\forall y \in U_0 \setminus N$ is sufficient.*

Proof. The strategy is the following. Let us assume we have a non-sufficient phase point x fixed for which the collision sequence has the structure (a, b, c) described above. We will show that x lies on a one-codimensional submanifold N. The advances (for a fixed neutral vector) are α, β and γ; furthermore arbitrary fixed non-collision time moments directly before and after collision with b are denoted by t_- and t_+, respectively. According to the possible geometrical position of the cylinders we distinguish three cases.

Case 1. The base subspaces for the first two collisions are transversal and non-orthogonal, thus $L_a \cap L_b = \{0\}$ and $L_a \not\perp L_b$.

We examine a neutral vector w at time moment t_+. By Lemma 2.7:

$$w_a = (\alpha - \beta)v_a^- + \beta v_a^+; \qquad w_b = \beta v_b^+. \tag{3.1}$$

Let us denote by J the linear mapping of $\mathbf{R^d}$ onto itself which is evaluated as the sum of the two orthogonal projections onto the subspaces L_a and L_b for any vector z, i.e. $Jz = z_a + z_b$. The transversality of the corresponding generator subspaces $A_a \cap A_b = \{0\}$ required for our model, which is equivalent to $L_a + L_b = \mathbf{R^d}$, ensures that J is indeed a linear bijection. Summing the two equations in (3.1) and applying J^{-1} we get:

$$w = \beta v^+ + (\alpha - \beta)J^{-1}(v_a^-)$$

On the other hand it is clear from (2.2) that for any neutral vector directly before collision with c: $w - \gamma v^+ \in A_c$, thus

$$(\beta - \gamma)v^+ + (\alpha - \beta)J^{-1}(v_a^-) \in A_c.$$

This equation however implies that unless two consecutive advances are equal (which is equivalent to sufficiency for our case, see the remark before proposition 3.1), one has:

$$\exists \lambda \in \mathbf{R}; \lambda \neq 0 : \quad v^+ + \lambda J^{-1}(v_a^-) \in A_c. \tag{3.2}$$

Let us now apply a purely configurational infinitesimal translation δq to the phase point of the trajectory segment at time t_+. Under the effect of all such translations, while v^+ remains the same, the velocity at t_- may change from v^- to $\bar{v}^-$ (more precisely, the velocity difference $v^- - \bar{v}^-$ moves on the surface of a $dim(L_b)$ dimensional sphere inside L_b that goes through the origin). Assume that the perturbed trajectory is not sufficient either, i.e. (3.2) holds for $\bar{v}^-$ with some $\bar{\lambda} \in \mathbf{R}$. This implies:

$$J^{-1}(\lambda v_a^- - \bar{\lambda}\bar{v}_a^-) \in A_c.$$

Now observe that by $L_a \cap L_b = \{0\}$ it is only true for vectors inside A_b that their images under J lie in L_a. This together with the required transversality of generator subspaces ($A_b \cap A_c = \{0\}$) and the bijectivity of J gives:

$$\lambda v_a^- = \bar{\lambda}\bar{v}_a^- \quad \Rightarrow \quad v_a^- \| \bar{v}_a^-. \tag{3.3}$$

We should distinguish two further subcases.
(i) $L_a \cap A_b = \{0\}$. The points of N are described by formula (3.3). However, under the effect of all perturbations δq the velocity difference projected onto L_a (i.e. $v_a^- - \bar{v}_a^-$) moves on a surface of an ellipsoid in the space L_a. Thus we get a contradiction with (3.3), x lies on a submanifold N of codimension at least 1.
(ii) $L_a \cap A_b \neq \{0\}$. Then by dynamics $P_{A_b}(v^-) = P_{A_b}(v^+)$, thus, as v^+ remains unchanged

$$P_{L_a \cap A_b}(v^-) = P_{L_a \cap A_b}(\bar{v}^-). \tag{3.4}$$

Now we shall see $N = N^{(1)} \cup N^{(2)}$ with both $N^{(1)}$ and $N^{(2)}$ one-codimensional. For the points $x \in N^{(1)}$ let

$$P_{L_a \cap A_b}(v^-) = P_{L_a \cap A_b}(v^+) = 0, \tag{3.5}$$

this submanifold is clearly (at least) one-codimensional. On the submanifold $N^{(2)}$ we assume that such a restriction of the velocity is not present. Then by (3.4) we see that (3.3) can only hold if $\lambda = \bar{\lambda}$, i.e. $v_a^- = \bar{v}_a^-$ (in fact, this is the formula that characterizes the points of $N^{(2)}$). Nevertheless v_a^- can only remain unchanged under the effect of all perturbations δq if $L_a \perp L_b$, something we excluded from case 1. As a perturbation moves away from $N^{(2)}$, it is definitely of codimension at least one.
Thus in case 1 it is indeed true that sufficiency may only fail locally on a submanifold N for which $codim(N) \geq 1$.

Remarks. Observe that our argument does not use any assumptions on the geometrical position of the cylinder C_c, geometrical restriction is only put on the cylinders of consecutive collisions a and b, and not on b and c. E.g. the case $L_b \perp L_c$ is also covered.

If $L_b \cap L_c \neq \{0\}$, the situation is even better. In addition to the characterization of non-sufficient points described above, by (2.2) it is straightforward to see that non-equality of the consecutive advances β and γ (i.e. non-sufficiency) can only hold if

$$P_{L_b \cap L_c}(v^+) = 0. \tag{3.6}$$

Thus for the submanifold N of non-sufficient points in fact $codim(N) \geq 2$ is valid, as we can find two transversal submanifolds N_1 and N_2 such that $N = N_1 \cap N_2$ holds and $codim(N_j) \geq 1$ (directions in configurational and velocity space, just like two velocity restrictions in orthogonal subspaces are always transversal).

Case 2. The subspaces L_a and L_b are not transversal, i.e. $L_a \cap L_b \neq \{0\}$.

Repeating the argument in the remarks after case 1 we get $P_{L_a \cap L_b}(v^-) = 0$ which gives for the submanifold of non-sufficient points $codim(N) \geq 1$ readily. It is however interesting to note that in many cases we automatically get $codim(N) \geq 2$:
(i) if $L_b \cap L_c = \{0\}$ and $L_b \not\perp L_c$; we are exactly in the situation discussed in the remarks above, thus $codim(N) \geq 2$;
(ii) if $L_b \cap L_c \neq \{0\}$, we have $N = N_1 \cap N_2$, where N_2 is defined by (3.6) and $P_{L_a \cap L_b}(v^-) = 0$ is valid for the points of N_1. These two are transversal, as applying all possible purely configurational translations δq at time t_-, v^- does not change (so we remain on N_1). On the other hand v_b^+ moves on a surface of a sphere of full dimension in L_b, thus (3.6) does not remain true for any such δq. We get $codim(N) \geq 2$.
(iii) for the case $L_b \perp L_c$ we are satisfied with $codim(N) \geq 1$.

Case 3. The two consecutive base subspaces are orthogonal, $L_a \perp L_b$. Observe that, from the viewpoint of Proposition 3.1, there is no need to consider this case separately, as by time direction change it is equivalent to one of the cases 1 or 2 (orthogonality of a and b together with b and c cannot occur, in that case we would have a shorter equivalent form). Nevertheless it is worth to get the explicit form of N as a condition on the velocity at time t_+ to prepare the proof of Proposition 3.2.

Observe that by the assumed $A_a \cap A_b = \{0\}$ the condition $L_a \perp L_b$ means $L_a = A_b$ and $L_b = A_a$. In case $L_b \cap L_c \neq \{0\}$ (if $L_a \cap L_c \neq \{0\}$, by equivalence, similarly) we get N explicitly as (3.6). If no such intersection is present, then the cylinders C_b and C_c are transversal in a strong sense, i.e. $L_b \cap L_c = \{0\}$; $A_b \cap L_c = \{0\}$ (this is the geometric condition that was required from the two cylinders in [16]). It is easy to see that under such geometrical conditions $(dim(L_b) =) dim(A_c) \geq 2$. From here on the velocity vector v and the neutral vectors w are always understood in time t_+. By orthogonality (2.3) gives a very simple form for any neutral vector:

$$w_a = \alpha v_a; \quad w_b = \beta v_b.$$

By a standard argument similar to the one used in case 1:

$$\exists \lambda \in \mathbf{R}, \lambda \neq 0 : \quad v_a + \lambda v_b \in A_c. \tag{3.7}$$

This formula is characteristic for the set of nonsufficient points N, for which $codim(N) \geq 1$. Indeed, let us assume that (3.7) holds for v and for any $\bar{v}$ which is parallel to $(v + \delta v)$

for some $\delta v \in A_c$ (we may assume $v \notin A_c$, otherwise we get codimension 2 readily). Let the corresponding constants in (3.7) be λ and $\bar{\lambda}$, respectively. The condition on both vectors implies:

$$(\bar{\lambda} - \lambda)v_b + (\bar{\lambda} - 1)\delta v_b \in A_c.$$

As we assumed $A_b \cap A_c = \{0\}$ there is definitely a possible choice of $\delta v \in A_c$ such that δv_b and v_b are lineary independent. This – by the strong transversality mentioned above – gives $\lambda = \bar{\lambda} = 1$, and looking at (3.7) we get $v \in A_c$ as a contradiction.

Thus we have seen $codim(N) \geq 1$ in all the three cases, the Proposition is proven. $\square$

It is time to fix the notion of richness we want to use for our model. We say that a trajectory segment is **rich** if its collision sequence contains
(i) either four consecutive archipelagos;
(ii) or three archipelagos at least one of which consists of more than one island.
(See, however Remark 2.6.) It is easy to see that such rich sequences, in any of their shortest possible equivalent forms, contain a subsequence of at least four consecutive islands (a, b, c, d) (with double occurrence of the same cylinder not excluded) such that $L_a \perp L_b$ and $L_c \perp L_d$ does not hold simultaneously. Observe furthermore that double connectedness of the collision sequence (already mentioned in subsection 2.3) would mean for our model the presence of four consecutive archipelagos; thus the notion of richness we use is weaker.

Proposition 3.2 *Assume $\Sigma = (a, b, c, d)$ is (an above described subsequence of) an arbitrary rich collision sequence. For any regular phase point $x \in M^0$ a bounded trajectory segment of which has collision sequence Σ there exist an open neighborhood $x \in U_0$ and a submanifold N of M such that (i) $codim(N) \geq 2$ and (ii) $\forall y \in U_0 \setminus N$ is sufficient.*

Proof. Now we shall think of x as a non-sufficient phase point with one particular collision sequence Σ, and show that x lies on the two-codimensional submanifold N. On our collision sequence $\Sigma = (a, b, c, d)$ we fix some arbitrary non-collision time moments t_- between collisions a and b, t_0 between b and c and t_+ between c and d. As for the rest of the notation we adopt the conventions of Proposition 3.1. Quantities without upper indices are calculated at time t_0. As before, we have to discuss several cases.

Case 1. $L_b \cap L_c \neq \{0\}$. As we want to consider (suitable subsequences of) rich sequences $L_a \perp L_b$ and $L_d \perp L_c$ cannot hold simultaneously. We may assume e.g. $L_b \not\perp L_a$. Then a quick reference to case 2 in the proof of Proposition 3.1 shows that $codim(N) \geq 2$ is ensured even by the subsegment (a, b, c) of Σ.

Case 2. $L_b \cap L_c = \{0\}$, $L_a \cap L_b = \{0\}$ and $L_c \cap L_d = \{0\}$. From Proposition 3.1 applied to two subsegments containing three islands we know that $N = N_1 \cap N_2$ (as we assumed that the sequence (a, b, c, d) has no shorter equivalent form, two consecutive orthogonality relations are not possible). All we have to prove is the transversality of the (at least 1 codimensional) submanifolds N_1 and N_2. As double orthogonality is not allowed there are two subcases. We shall discuss (i) in full detail and give some hint related to (ii).

(i) $L_a \perp L_b$ and $L_c \not\perp L_d$ (or the other way round). N_1 is defined as a velocity condition in t_0 (by equation (3.7)), as for N_2 we have (by case 1 from Proposition 3.1) two possibilities. If there exists a purely configurational translation transversal to N_2, the required transversality is trivial. Otherwise the points of N_2 are defined by the velocity restriction

$$(P_{L_d \cap A_c}(v) =) P_{L_d \cap A_c}(v^+) = 0. \tag{3.8}$$

By $L_c \not\perp L_d$ and $L_c \cap L_d = \{0\}$, there definitely exists a vector $0 \neq \delta v \in A_c$ such that $\delta v \perp L_d \cap A_c$. Apply the velocity perturbation $v \to \bar{v} = (v + \delta v)^o$ (here and from here on given any non-zero vector z, $(z)^o$ stands for the unit vector which is a positive multiple of z). Tangentiality to N_2 is obvious, thus we are done if we show transversality to N_1. An adaptation of the argument from Case 3 in Proposition 3.1 clearly works unless v_b and δv_b are lineary dependent for any choice of δv. If such an accidental situation comes about then the (generally at least two dimensional) velocity component v_b should be restricted to a line determined by the geometrical position of our cylinders. This gives $codim(N) \geq 2$ together with (3.8) above by the transversality of A_c and $A_a = L_b$.

(ii) $L_a \not\perp L_b$ and $L_c \not\perp L_d$. There are two possibilities for both N_1 and N_2 (see formulas 3.3 and 3.5). In any of these possibilities the transversality of the two manifolds is not difficult to show and we leave the details to the reader. What is of key importance is the pairwise transversality of generator subspaces (see also the end of section 3 from [16]).

Case 3. In addition to $L_b \cap L_c = \{0\}$ e.g. $L_c \cap L_d \neq \{0\}$. The required $codim(N) \geq 2$ we get from the collision sequence (b, c, d) readily unless $L_b \perp L_c$. If such an orthogonality is present, we get $N = N_1 \cap N_2$. N_1 is defined as a condition on the velocity v^- at time t_- (see formula 3.7). For the phase points belonging to N_2 the condition $P_{L_c \cap L_d}(v^+) = 0$ holds at time t_+. This later condition can be surely spoilt by applying purely configurational translations at time t_-. Under the effect of such translations, however, the velocity v^- does not change. Thus we have found δq tangential to N_1 and transversal to N_2, the required transversality of N_1 and N_2 is proven.

We have seen $codim(N) \geq 2$ in all possible cases, thus the Proposition is proven. □

3.2 Dynamical-Topological Considerations and finishing the proof

Let us denote by M_p^0 the set of regular and poor points on our phase space. The aim of the dynamical-topological part would be to prove that M_p^0 is a slim subset; there may exist, however, some trivial 1-codimensional submanifolds of non-sufficient points. Therefore we should restrict our considerations to a subset $M^{\#} \subset M$. Just like in [21] and [16], we define the complement of $M^{\#}$, $M \setminus M^{\#}$ as the set of points which have trajectories without any collision in at least one of the non-trivial sub-billiards of our original billiard (as to the notion of a sub-billiard see subsection 2.1). Applying the considerations of Appendix 2 from [21] to our model, we see that $M^{\#}$ has a finite number of connected components. In the rest of the subsection we will prove that any of these components belongs to one ergodic component. Then the proof of Theorem 2.4 is finished exactly the same way as it was done in Lemma A.2.3 in [21] by connecting the components of $M^\#$ with bundles of orbits of positive measure.

Let us first examine the set of poor phase points a bit closer. We have the natural partition $M_p^0 \cap M^\# = M_1^0 \cup M_2^0 \cup M_{3-}^0$ where the lower index i refers to the number of archipelagos of the collision sequence of $x \in M_i^0$. We use M_{3-}^0 as not all collision sequences with three archipelagos are poor (more precisely, for $x \in M_{3-}^0$ there are three archipelagos, each containing only one island). However, it is useful to further subdivide M_2^0 into M_{2-}^0 and M_{2+}^0, where the points $x \in M_{2-}^0$ are characterized by the fact that both archipelagos contain only one island. This partition is useful as we may treat the points of $M_1^0 \cup M_{2-}^0$ and $M_{2+}^0 \cup M_{3-}^0$ separately. Indeed, observe that for the latter one the statement of Proposition 3.1 holds, a fact we will strongly use in the proof of Lemma 3.4 below.

We introduce the notation M_{ns} for the set of non-sufficient points of our billiard. The essence of the dynamical-topological part is the following Proposition:

Proposition 3.3 *The set of points $M_p^0 \cap M_{ns} \cap M^\#$ is slim.*

Proof of this Proposition, closely following section 4 from [16], consists of several steps according to the partition of the 'poor' set described above. As the slimness of $M_1^0 \cap M_{2-}^0$ is an easy adaptation of former ball avoiding theorems, we just give a sketch with the suitable references; while the rest of the proof is discussed in a more self-contained way (Lemma 3.4).

The slimness of M_1^0 we get by a trivial adaptation of the classical ball avoiding Theorem from [7] (as for a detailed description see [23] as well). Indeed, for a phase point $x \in M_1^0$ the entire trajectory avoids at least one cylinder (otherwise there would be more archipelagos). On the other hand, the dynamics for such a phase point is determined by a non-trivial sub-billiard. This sub-billiard is, however, modulo almost periodic motion, either dispersive or the product of two dispersive billiards. In the former case the classical strong ball avoiding theorem (see [23]), in the latter a simple generalization of it (cf. Lemma 4.6 from [21]) does the job.

As for the points $x \in M_{2-}^0$ the island structure of the trajectory is (a, b) with $L_a \not\angle L_b$. We can repeat the proof of Lemma 4.1 from [16] word by word; problem only arises at the analogue of Sublemma 4.5 where the geometry of the cylinders is used. However, it is possible to refer to Proposition 3.10 from [19] where we find the needed generalization readily. Thus we have seen the slimness of $M_1^0 \cup M_{2-}^0$. It is enough to show something weaker for the rest of the 'poor' set to finish the proof of the Proposition:

Lemma 3.4 *The subsets $N_0 = M_{3-}^0 \cap M_{ns}$ and $N_0' = M_{2+}^0 \cap M_{ns}$ are both slim.*

Proof of this Lemma again relies on [16] (it is an adaptation of the proof of Lemma 4.2). Let us start with N_0. We may denote the island sequence of the examined point $x \in N_0$ by (a, b, c) and fix some non-collision time moments, t_- between islands a and b and 0 between islands b and c (N.B. these are cutting time moments). We note that no orthogonality is present, furthermore it can be assumed that either (i) $L_a \cap L_b = \{0\}$ or (ii) $L_b \cap L_c \neq \{0\}$ (otherwise reverse time direction). By Proposition 3.1 we know that there is a small open neighborhood U of x and a codimension-one submanifold N such that $U \cap N_0 \subset N$. The explicit form of N is defined by cases 1 and 2 from Proposition 3.1 for the possibilities (i) and (ii) above, respectively. We define the set

$$F_+ := \{x \in N : \text{the positive semi-trajectory of } x \text{ has proper collisions only with } C_c\}.$$

As $U \cap N_0$ is a subset of F_+, to finish the proof it is enough to show that the closed set F_+ contains no open disk D in N. Assume the contrary. Now – following [16] – we have to find manifolds $\gamma(y)$ that go through the points $y \in D$ transversally to N (and thus to D). We always think of these manifolds in an infinitesimal sense, i.e. with inner radius smaller than a fixed small ϵ. By their help the set

$$\bar{D} := \cup_{y \in D} \gamma(y)$$

is constructed, the positive semi-trajectory of which – although does not avoid the original cylinder C_b necessarily – surely does avoid the modified cylinder we obtain by shrinking C_b by ϵ. Just like in [16], a quick reference to the weak ball-avoiding theorem, applied to the sub-billiard defined with lone cylinder C_c (see [7, 23]) gives $\mu(\bar{D}) = 0$, a contradiction.

The construction of $\gamma(y)$ should be done, due to the fact that we have various cases, in a way a bit different from the one in [16]. From here on $y = (q, v)$ is an arbitrary fixed point in D. First we define the so-called pseudo-unstable manifolds:

$$\gamma_0^u(y) = \{z = (q', v') : v' = v, \ q' - q \in A_c, \|q' - q\| < \epsilon/2\}$$

$$\gamma_e^u(y) = \{z : dist(y, z) < \epsilon/2; \ dist((S_c)^t z, (S_c)^t y) \to 0 \text{ exponentially fast as } t \to +\infty\}$$

And then construct $\gamma(y)$ as:

$$\gamma(y) = \cup_{z \in \gamma_0^u(y)} \gamma_e^u(z) = \cup_{z \in \gamma_e^u(y)} \gamma_0^u(z).$$

Here the notation $(S_c)^t$ stands for the dynamics determined by the sub-billiard of the original system the configuration space of which we get by cutting out only the cylinder C_c from $\mathbf{T^d}$. Observe that $\gamma_e^u(y)$ is the stable manifold through y for this sub-billiard, while $\gamma_0^u(y)$ is analogous to the manifold that was used as $\gamma(y)$ in [16]. By the explicit form of the submanifolds N (cases 1 and 2 in Proposition 3.1, respectively) we know that these manifolds γ are indeed transversal to N. More precisely, in case (ii) N is described by (3.6) and γ_e^u is always transversal to it. In case (i) there are two possibilities. If the points of N are defined by (3.3), then it follows from the pairwise transversality of generator spaces that γ_0^u and N are always transversal (see Lemma 4.6 from [16]). Otherwise N is given by formula (3.5), and, as a further consequence of the pairwise

transversal generator spaces, γ_e^u does the job.
It is also straightforward that the set $\bar{D}$ avoids the shrunk cylinder.

As to the points of N_0' we choose 0 as the cutting time moment between the two consecutive archipelagos. On the positive semi-trajectory there are collisions with two cylinders C_a and C_b such that $L_a \perp L_b$. The last collision before 0 is with some C_c, where $L_c \not\perp L_a$ and $L_c \not\perp L_b$. Before that there may occur some further preceding collisions but that makes no difference for us. In our local analysis we again have $U \cap N_0' \subset N$ where N is now described by case 3 from Proposition 3.1. We define F_+ as the set of points on N which do not collide with any cylinder except C_a and C_b in the future. Submanifolds for which transversality to N is to be shown (they trivially do avoid the shrunk cylinder):

$$\gamma(y) = \{z : dist(y,z) < \epsilon;\ dist((S_{a,b})^t z, (S_{a,b})^t y) \to 0 \text{ exponentially fast as } t \to +\infty\}.$$

Here $(S_{a,b})^t$ is the sub-billiard dynamics we get by cutting out only the two orthogonal cylinders C_a and C_b. We note that by orthogonality this dynamics is a product of two dispersive billiard systems, and, as a consequence, the manifolds $\gamma(y)$ are products of the stable manifolds for these two dispersive billiards. The transversality of $\gamma(y)$ and N is easily seen if $L_a \cap L_c \neq \{0\}$ or $L_b \cap L_c \neq \{0\}$, as in that case N is described by (3.6). If on the contrary no such intersection is present, the required transversality may be shown indirectly. Indeed, points of N are described by the (at least) one-codimensional velocity restriction (3.7). Now it is easy to see that if we combine $\gamma(y)$ with the infinitesimal perturbation (3.9) below and project onto the velocity sphere, all points of $\mathbf{S^{d-1}}$ in a small open neighborhood of v can be covered. However, by the explicit form of (3.7) the velocity translation

$$v = v_a + v_b \ \to\ \bar{v} = ((1-\epsilon)v_a + (1+\epsilon)v_b)^o \tag{3.9}$$

is clearly tangential to N.
The demonstration of our Lemma is complete. □ Hence the Proposition is proven □.

Our last task is to consider singular trajectories. More precisely, we should prove the analogues of Propositions 3.2 and 3.3 for points of M^1 on the one hand and verify the Chernov-Sinai Ansatz on the other hand. Following tradition ([10, 21, 16]) we state Proposition 3.5 below from which both the Ansatz and the analogous Lemmas follow. In our setting *eventually simple phase points* are points of singular reflections $x \in \mathcal{SR}^+$ for which (i) the positive semi-trajectory $S^{(0,\infty)}x$ is regular and (ii) there exists a time moment $t > 0$ such that the collision sequence of $S^{[t,\infty)}$ contains only one archipelago.

Proposition 3.5 *For every cell R of maximal dimension $2d-3$ in $\mathcal{SR}^+$, the set $R_{es} \subset R$ of eventually simple points can be covered by countably many closed zero-subsets (with respect to the induced measure μ_C) of C.*

Proof of this Proposition, just like as it was with the analogous statements in [16, 21], is an adaptation of the proof of Main Theorem 6.1 in [10]. Let us denote one particular eventually simple collision sequence (containing only one archipelago) by Σ. There are two possible cases, either $\Sigma \sim (a)$ or $\Sigma \sim (a, b)$ with $L_a \perp L_b$. For both of them there is definitely at least one cylinder C_c avoided by the semi-trajectory $S^{[t,\infty)}$. In other words our eventually simple phase point belongs to a ball avoiding set F_+ (the definitions of y_0, $U(y_0)$, F_+, F'_+ are taken from the above references). We also define the pseudo-stable invariant manifolds $\gamma_0^{sp}(y)$ and $\gamma_e^s(y)$ according to the literature: if $\Sigma \sim (a)$, we have $dim(\gamma_0^{sp}(y)) = dim(A_a)$ and $dim(\gamma_e^s(y)) = dim(L_a) - 1$; while for $\Sigma \sim (a, b)$ we get $dim(\gamma_0^{sp}(y)) = 1$ and $dim(\gamma_e^s(y)) = dim(L_a) + dim(L_b) - 2$. However in both cases we arrive, just like in the references, at the $d-1$-dimensional $\gamma_g^s(y)$, which is a concave orthogonal manifold. The key point is again that backward images of concave orthogonal manifolds are always transversal to the set of singular reflections (as it is true by sublemma 4.2 in [8]). This together with weak ball avoiding gives an indirect proof of the Proposition (similarly to the arguments in Lemma 3.4). □

Propositions 3.2, 3.3 and 3.5 altogether imply that any connected component of the set $M^\#$ belongs to one ergodic component. As already mentioned at the first paragraph of the subsection, a reference to Lemma A.2.3 from [21] finishes the proof of the required global ergodicity.

4 Theorem 2.5 – three cylinders

In this section our aim is to prove, following the steps (H1)-(H2) mentioned in subsection 2.2, the hyperbolicity of a billiard with three cylinders which satisfy the transitivity condition of conjecture 2.1. Observe that the transitivity condition may only hold if there is at most one orthogonal pair (i.e. a pair i, j with $L_i \perp L_j$) among the three cylinders (otherwise there would be a non-trivial orthogonal splitting present, see Remark 2.3). As a cosequence the archipelago structure is unique for any collision sequence. There are two possible types of archipelagos: an archipelago either contains one single island, or it is a maximal sequence of alternating islands of two orthogonal cylinders.
In symbolic collision sequences we refer to the three cylinders C_1, C_2, C_3 simply by the numbers 1, 2 and 3. In all our notation we follow section 3, there is only one difference: to simplify the calculation of neutral subspaces in subsection 4.1, we adopt a convention from [10]; we fix the advance of one particular (central) collision zero. That way sufficiency of the segment is equivalent to the triviality of the neutral subspace, $\mathcal{N} = \{0\}$.

Of course, all trivially sufficient trajectories (in the sense of Remark 2.6) are considered to be rich. Besides, the specific **notion of richness** we use for our model is exactly the double connectedness already mentioned in subsection 2.3: a phase point x is rich if there exists a cutting time moment t on its entire trajectory, such that the symbolic collision sequences for both segments $S^{(-\infty,t]}x$ and $S^{[t,\infty)}x$ are connected (i.e. they contain enough cylinders to generate a transitive action).
The essence of our geometric-algebraic considerations (subsection 4.1) is to prove

Proposition 4.1 *Assume $x \in M^0$ is an arbitrary rich and regular phase point. There* *exist an open neighborhood $x \in U_0$ and a submanifold N of M such that (i) codim$(N) \geq 1$ and (ii) $\forall y \in U_0 \setminus N$ is sufficient.*

As for the dynamical-topological part our key statement is

Proposition 4.2 *Let us denote by M_p^0 the set of regular and poor phase points. The set M_p^0 is of zero μ-measure.*

Actually in subsection 4.2 we prove a stronger statement (Lemma 4.4) which implies the Proposition above. We finish the discussion of Theorem 2.5 in the same subsection.

4.1 Geometric-Algebraic Considerations

In the case of three cylinders one easily gets a classification of all possible rich collision sequences. Indeed, a rich collision sequence necessarily contains a subsequence Σ which has (up to equivalence) one of the forms below:
(a) $\Sigma \sim (1, (2-3)_{\geq 3}, 1)$ (here $(2-3)_{\geq 3}$ stands for a sequence of the cylinders C_2 and C_3 which contains at least three consecutive collisions in the shortest equivalent form);
(b) $\Sigma \sim ((2-3)_{\geq 3}, 1, 3)$;
(c) $\Sigma \sim (a, b, c)$ with $A_a \cap A_b = A_a \cap A_c = A_b \cap A_c = \{0\}$ (the possibility $a = c$ is not excluded);
(d) $\Sigma \sim (1, 2, 3, 1)$ with $A_1 \cap A_2 = A_1 \cap A_3 = \{0\}$ but $A_2 \cap A_3 \neq \{0\}$;
(e) $\Sigma \sim (1, 2, 3, 1, 2)$ with $A_1 \cap A_3 \neq \{0\}$
A few hints are in order to show how one can get such a classification. If we disregard the trivially sufficient cases mentioned in Remark 2.6, any rich sequence should contain at least four consecutive islands. Now there are two possibilities for the shortest possible equivalent forms: either there is a subsegment of type $(2-3)_{\geq 3}$, or our rich sequence is a finite part of the infinite sequence $(...1, 2, 3, 1, 2, 3, ...)$. In the presence of $(2-3)_{\geq 3}$, a transversality relation $A_2 \cap A_3 = \{0\}$ leads to case (c). Without this transversality the segment $(2-3)_{\geq 3}$ is not even connected and continuing it to a rich sequence gives one of the cases (a) or (b). Now let us examine $\Sigma \sim (1, 2, 3, 1, ...)$; we may assume that there is no orthogonal pair among the three cylinders (indeed, otherwise our sequence is either not rich or equivalent to one of the above discussed cases). Assume first that $A_1 \cap A_2 = \{0\}$ and $A_1 \cap A_3 = \{0\}$ simultaneously, we arrive either at (c) or at (d). We note that the first four islands are enough for richness if and only if both of these transversalities hold (remember orthogonality is excluded). The other possibility is that at least one of these transversality relations does not hold and one or two additional islands should be present. If $A_1 \cap A_3 \neq \{0\}$ we are automatically at case (e). Otherwise $A_1 \cap A_3 = \{0\}$ and $A_1 \cap A_2 \neq \{0\}$, the rich sequence contains six islands and it has a subsequence $(2, 3, 1, 2, 3)$ which is described by case (e).

Now our task is to describe the one-codimensional submanifolds N for all the above five cases. We shall discuss the arguments for cases (a) and (e) in full detail; as to the rest we just give some hints and leave the proof to the reader.

Case (a). Non-collision time moments directly after the first and before the second collision with C_1 are denoted by t_- and t_+, while the advances of these two collisions are α and α', respectively. We shall get $N = N^{(1)} \cup N^{(2)}$ with both $N^{(1)}$ and $N^{(2)}$ one-codimensional. Observe that we may exclude the possibility $L_2 \perp L_3$, otherwise we would have a shorter equivalent form.

To get $N^{(1)}$ consider the trajectory segment $S^{[t_-,t_+]}x$. Time evolution in this shorter time interval is determined by the sub-billiard dynamics defined by cutting out only two cylinders C_2 and C_3 (this can be understood on the torus $\mathbf{T}^{\mathbf{d}'}$ where $d' = dim(L_2 + L_3)$, together with almost periodic motion in the directions of $A_2 \cap A_3$). From Proposition 3.1 we however know that, in a small open neighborhood of x phase points non-sufficient with respect to this sub-dynamics in the time interval $[t_-, t_+]$ lie in a one-codimensional submanifold $N^{(1)}$ (remember $L_2 \not\perp L_3$!).

Consider now on the contrary phase points which do not lie on $N^{(1)}$, i.e. which are sufficient with respect to the above mentioned subdynamics. Among them we would like to characterize those non-sufficient in the whole time interval $[a, b]$ for the full billiard dynamics. We shall show below that such points form a one-codimensional submanifold $N^{(2)}$. The advances of all the collisions with C_2 and C_3 are equal, so we may use the convention mentioned in the beginning of section 4 and fix the advances for all these central collisions to be zero. Observe that for any neutral vector calculated at time moments t_- and t_+: $w^- = w^+ = w \in A_2 \cap A_3$. Moreover, neutrality with respect to the first and the second collisions with C_1 implies (by (2.2))

$$\alpha v_1^- = w_1^- = w_1^+ = \alpha' v_1^+.$$

As $\alpha = 0$ or $\alpha' = 0$ would mean sufficiency we may conclude that

$$v_1^- \| v_1^+. \tag{4.1}$$

Apply now all possible purely configurational translations δq at time moment t_-; v^- does not change while the velocity difference $v^- - v^+$ moves on a surface of a sphere of highest possible dimension in $L_2 + L_3$. Following the argument at the end of case 1 from Proposition 3.1, we see that the perturbations δq definitely give a direction transversal to the submanifold $N^{(2)}$ of non-sufficient points. ($L_1 \perp (L_2 + L_3)$ is not possible, in that case the transitivity condition would not hold for our three cylinders). Thus $codim(N^{(2)}) \geq 1$.

The discussion of **Case (b)** is analogous to Case (a) and we leave it to the reader (see also cases 9 and 11 in section 4 of [10]). In **Case (c)** we may directly refer to Proposition 3.1 to see that the submanifold N is indeed (at least) one-codimensional.

As to **Case (d)** the key is again reference to Proposition 3.1 with some adaptation. However, a few hints are in order. We may assume that (i) any two of the three base subspaces are transversal (otherwise we have sufficiency apart from codimension 1 by case 2, Proposition 3.1); (ii) there is no orthogonal pair among the three cylinders (otherwise our sequence would not be rich, and, continuing it in any way to a rich sequence would lead to one of the cases (a) or (b)). Now the whole discussion is analogous to case 1 from Proposition 3.1 (hint: calculate w_1 and w_3 by Lemma 2.7 for a neutral vector w directly after the island with C_2 and apply configurational translations directly before that island).

In all the above cases we could describe the one-codimensional manifold N with the help of former methods (in cases (c) and (d) by techniques from subsection 3.1, in cases (a) and (b) by adaptation of the proofs from [10] to the cylindric billiard setting). The situation is however quite different with **Case (e)**.

Lemma 4.3 *Consider a phase point x with collision sequence $\Sigma \sim (1,2,3,1,2)$, where $A_1 \cap A_3 \neq \{0\}$ (i.e. case (e)). The statement of Proposition 4.1 is true for x.*

Proof. We may assume that neither orthogonality, nor intersection is present for any pair of base subspaces. Indeed, in case $L_i \perp L_j$ $(i \neq j)$ our collision sequence would be equivalent to any of the above cases (a) or (b). On the other hand if $L_i \cap L_j \neq \{0\}$ for some $i \neq j$, we would have sufficiency in a sub-billiard for the subsegment (i,j) outside a 1-codimensional manifold $N^{(1)}$, and a reference to the arguments in cases (a) or (b) would give $N = N^{(1)} \cup N^{(2)}$.

We choose 3 as our central collision and fix the advance as zero. Notation for the other advances:
α' for the first collision with C_1 and α for the second;
β for the first collision with C_2 and β' for the second.
We exclude the possibilities $\alpha = 0$ and $\beta = 0$, in these cases we would have sufficiency in a sub-billiard for one of the sequences $(3,1)$ or $(2,3)$ and could repeat the argument from case (a). The non-collision time moments we fix are:
t_* after the first 1 and before the first 2;
t_- after the first 2 and before 3;
t_+ after 3 and before the second 1;
$t_\#$ after the second 1 and before the second 2.
Depending on the geometrical position of the cylinders we distinguish two further subcases.

Subcase (e1). $dim(L_1 \cap A_2) \leq 1$, thus the two subspaces are either transversal or the intersection $L_1 \cap A_2$ is a line. By the convention of zero advance at the central collision we know that for any neutral vector at time moments t_- and t_+, $w^+ = w^- = w \in A_3$. Moreover, by applying (2.3) to our collision sequence we get the set of equations:

$$\begin{array}{rcccl} \alpha v_1^+ & = & w_1 & = & (\alpha' - \beta)v_1^* + \beta v_1^- \\ \beta v_2^- & = & w_2 & = & (\beta' - \alpha)v_2^\# + \alpha v_2^+. \end{array} \tag{4.2}$$

The first equation implies, that for non-sufficient points:

$$\exists \lambda \in \mathbf{R}; \lambda \neq 0 : \quad (v_1^+ - \lambda v_1^-) \,\|\, v_1^* \tag{4.3}$$

while from the second equation in (4.2) we get that *with the same* λ:

$$(\lambda v_2^- - v_2^+)\,\|v_2^\#. \tag{4.4}$$

Now assume for a while that neither $v_2^-\|v_2^+$ nor $v_1^-\|v_1^+$ and apply a purely configurational translation at time moment t_- for which $\delta q \in A_1 \cap A_3$. Under the effect of such a perturbation:
(i) *none of the velocity vectors v^-, v^+ and $v^\#$ changes.* Thus if the perturbed phase points remain non-sufficient, as the vector quantities in (4.4) do not change, equation (4.3) holds throughout the perturbation with the original constant λ (remember that parallelity is excluded).
(ii) On the other hand as $A_1 \cap A_2 \cap A_3 = \{0\}$ (otherwise our cylinders could not satisfy the transitivity condition), under the effect of the above perturbation *the velocity vector at time moment t_* changes*,

$$v^* \;\rightarrow\; \bar{v}^* \tag{4.5}$$

where the velocity difference $v^* - \bar{v}^*$ moves on an arc of a circle that goes through the origin in L_2. (More precisely, if $dim(A_1 \cap A_3) = 1$, it might be possible that for such a perturbation $\delta q \in A_1 \cap A_3$ the velocity v^* does not change. However in that case $v_2^- \in ((A_1 \cap A_3) + A_2)$. Thus the (at least) two dimensional velocity component v_2^- is restricted to a line, which means a one-codimensional restriction for our phase point.)

Now we finish our argument similarly to case 1 from Proposition 3.1. Observe that by (i), as the left hand side of (4.3) is the same for v^* and $\bar{v}^*$:

$$v_1^*\|\bar{v}_1^*. \tag{4.6}$$

If $L_1 \cap A_2 = \{0\}$, then if we project orthogonally the arc of the perturbation (4.5) onto L_1 it remains an ellipse, thus (4.6) cannot hold. If $L_1 \cap A_2 \neq \{0\}$, then by the nature of dynamics

$$P_{L_1\cap A_2}(v^*) = P_{L_1\cap A_2}(v^-) = P_{L_1\cap A_2}(\bar{v}^*).$$

If $P_{L_1\cap A_2}(v^-) \neq 0$ (otherwise we have velocity restriction giving $codim(N) \geq 1$ itself), we know that (4.6) can only hold if $v_1^* = \bar{v}_1^*$, which is impossible as in subcase (e1) we assumed $dim(L_1 \cap A_2) \leq 1$.

Finally let us say a few words about the points for which e.g. $v_1^-\|v_1^+$. Apply all possible purely configurational translations at t_-; v^- does not change so for the perturbed velocity at time t_+ we expect $\bar{v}_1^+\|v_1^+$. However $\bar{v}^+ - v^+$ moves on a surface of a sphere of full dimension inside L_3, and, after projecting orthogonally onto L_1, we get (by $L_1 \not\perp L_3$): $codim N \geq 1$ exactly the same way as at the end of case 1 from Proposition 3.1. (Observe that the condition $dim(L_1 \cap A_2) \leq 1$ of subcase (e1) was not used in the analysis of this parallelity.)

Subcase (e2). $dim(L_1 \cap A_2) \geq 2$. At first we discuss what we can do if $(L_1 \cap A_2) \not\perp L_3$ (in other words if $L_1 \cap A_2 \not\subset A_3$). By the nature of dynamics for any neutral vector w we get the following series of equations:

$$P_{L_1\cap A_2}(w) = P_{L_1\cap A_2}(w^*) = \alpha' P_{L_1\cap A_2}(v^*) = \alpha' P_{L_1\cap A_2}(v^-).$$

Here w^* is the value of the neutral vector w at time moment t_*. On the other hand trivially:

$$P_{L_1 \cap A_2}(w) = \alpha P_{L_1 \cap A_2}(v^+).$$

Thus if the point is not sufficient:

$$P_{L_1 \cap A_2}(v^-) \| P_{L_1 \cap A_2}(v^+).$$

Apply all possible configurational translations at t_- to conclude, exactly the same way as with the parallelity discussed at the end of subcase (e1), that $codim(N) \geq 1$.

Problem only arises with the above argument if $L_1 \cap A_2 \subset A_3$, a possibility we discuss in an indirect way. First of all observe that in such a case there are a couple of generalizations for the main argument of subcase (e1). By $L_1 \cap A_2 \subset A_3$ trivially $A_2 \cap A_3 \neq \{0\}$. Thus we may apply the whole discussion word by word for the reverse directed sequence with $1 \leftrightarrow 2$, (in that case we would e.g. apply $\delta q \in A_2 \cap A_3$ at time t_+). It is also true that the scheme works for higher dimensional $L_1 \cap A_2$ as well whenever

$$dim(L_1 \cap A_2) \leq dim(A_1 \cap A_3). \tag{4.7}$$

Indeed, following the argument of subcase (e1), after application of $\delta q \in A_1 \cap A_3$ at time moment t_-, the dimension of the sphere on the surface of which the velocity difference $v^* - \bar{v}^*$ moves inside L_2 is $min(dim(A_1 \cap A_3), dim(L_2) - 1)$ (otherwise we would have a restriction on the velocity v^-, which would mean $codim(N) \geq 1$ itself). Problem only arises if there are too many orthogonal directions between L_1 and L_2, in other words if (4.7) does not hold (otherwise the effect of our perturbation is 'visible' at the orthogonal projection of our velocity difference onto L_1.)

Assume now that $dim(L_1 \cap A_2) = k \geq 1$. However, from $L_1 \cap A_2 \subset A_3$ trivially follows that $dim(A_2 \cap A_3) \geq k$, thus by the above considerations (now applied with $1 \leftrightarrow 2$) there is no problem if

$$dim(L_1 \cap A_2) \geq dim(L_2 \cap A_1). \tag{4.8}$$

Now we are ready with our indirect proof with one more reference to the $1 \leftrightarrow 2$ symmetry as either (4.8) or the opposite inequality surely holds.

The demonstration of the Lemma is ready. $\square$. Hence the Proposition. $\square$.

We close this subsection with an **example**. The strange geometric position of subspaces discussed in subcase (e2) above might seem quite unnatural (such a thing does not happen in hard ball systems, see also the appendix). One can however easily find an example from a well studied category, namely an "orthogonal cylindric billiard" (see [21] as basic reference for this example). In orthogonal cylindric billiards the base subspaces for the cylinders are given by subsets of a fixed orthonormal basis in $\mathbf{R^d}$. We define our billiard on $\mathbf{T^8}$, the base subspaces are

$$K_1 = \{1, 2, 3, 4\};\ K_2 = \{3, 4, 5, 6\};\ K_3 = \{6, 7, 8\}.$$

For this model $dim(L_1 \cap A_2) = 2$ and $L_1 \cap A_2 \subset A_3$.

4.2 Dynamical-Topological Considerations and finishing the proof

Proposition 4.2 is a direct consequence of the following two Lemmas:

Lemma 4.4 *Denote by M_e^0 the set of those regular phase points x for which there exists a time moment t_0 such that the segment $S^{[t_0,\infty)}x$ is not connected. $\mu(M_e^0) = 0$.*

Proof of this Lemma, as it might not be surprising for the reader, is again weak ball-avoiding ([7, 23]). Without loss of generality we may restrict our attention to points $x \in M_e^0$ for which $S^{[0,\infty)}x$ is not connected. Now by the orthogonal splitting (see remark 2.3) the dynamics for this semi-trajectory is determined by a product dynamics of K-system(s) and almost periodic motion (the 'K-factors' in the product dynamics possess the K-property as they are either dispersive billiards or the results of section 3 apply to them). On the other hand the positive semi-trajectory avoids one of the cylinders, e.g. C_3. However, for any of the product dynamics above weak ball-avoiding theorems can be applied, thus indeed $\mu(M_e^0) = 0$. □

Lemma 4.5 *Any symbolic collision sequence that contains three consecutive connected subsequences is rich.*

Proof. There are finite time moments $t_1 < t_3$ such that all the segments $(-\infty, t_1]$; $[t_1, t_3]$ and $[t_3, +\infty)$ are connected. It may happen that the sequence is rich in the sense of Remark 2.6 and than we are done. Otherwise it cannot happen that a connected collision sequence is contained in one archipelago. As a consequence there exists a cutting time moment t_2 with $t_1 \leq t_2 \leq t_3$. As both $(-\infty, t_2]$ and $[t_2, +\infty)$ are connected and t_2 is cutting, the whole sequence is doubly connected, thus rich (note that such a property is invariant under equivalence transformations). □

Propositions 4.1 and 4.2 together with the local hyperbolicity theorem (see subsection 2.2) give the proof of Theorem 2.5. Full hyperbolicity (the fact that the Lyapunov exponents are nonzero almost everywhere) implies, on the basis of Katok-Strelcyn theory ([6]), that the system has at most a countable number of ergodic components, each of positive measure. Moreover, on any of these ergodic components the dynamics is K-mixing (see e.g. [13]) and possesses the Bernoulli-property (see [3, 12]).

Concluding Remarks. Of course the natural question that arises how we could generalize the results of Theorems 2.4 and 2.5 further. If we increase the number of cylinders, we encounter more and more difficulty in the *geometric-algebraic considerations* as the number of collision sequences to be studied (with, moreover, different possible geometric positions of the cylinders) is much higher. This makes things especially complicated if we want to prove ergodicity, thus we should 'gain' two codimensions from the equations on the neutral subspaces. In a general setting (just like in our Proposition 3.2) one may hope to get $codim(N) \geq 2$ by $N = N_1 \cap N_2$ with N_1 and N_2 transversal and both one-codimensional. However, as the number of possible collision sequences increases, there are more and more possibilities for the manifolds N_1 and N_2 above, in many cases obtained in quite an implicit way (see Lemma 4.3), thus proving transversality for all cases seems very difficult, if not impossible (even in the setting of Theorem 2.4 one has to consider several possibilities, see Proposition 3.2). I think that proving the ergodicity part of conjecture 2.1 is at the present level much far away; we might not even guess what the suitable notion of richness for the proof of ergodicity could be. However for *proving hyperbolicity* even in a general setting *the notion of richness (double connectedness) we use in section* 4 *seems to be enough.* Though this task seems to be highly nontrivial at the geometric-algebraic part, we hope remarkable progress related to it ([14]).

As to the *dynamical-topological considerations*, although results from the literature were quite easily adapted to our setting, if we increase the number of cylinders this does not remain true. The main problem is that ball avoiding theorems, in their classical form, are *of inductive nature*, i.e. they rely on the K-property of smaller subsystems. For Hard Ball Systems a new type of ball avoiding theorem is used in [18], which is not inductive in its nature. This method, however, uses special symmetries of hard balls, thus its adaptation to the general cylindric billiard setting seems impossible.

Altogether we can say that if we want to discuss ergodic and hyperbolic properties for all cylindric billiards, i.e. to prove conjecture 2.1 in its full generality, a kind of breakthrough would be needed. Nevertheless the simple Lemma 2.7 could be a good starting point (for the geometric-algebraic considerations) even in such a general setting. We could say that *this Lemma 'saves' as much of the connected path formula* – which uses, e.g., the conservation of momenta in hard ball systems – *for the cylindric billiard setup as possible* (see also the Appendix). Probably the most remarkable consequence of our results is that they prove chaotic/ ergodic properties for full subclasses of cylindric billiards. Thus *we may get more convinced about the validity of conjecture* 2.1. In other words there is increasing evidence that the Boltzmann-Sinai Ergodic Hypothesis is true because Hard Ball Systems belong to the class of cylindric billiards that satisfy the transitivity condition.

Appendix: Some remarks on Hard Ball Systems

The most studied and physically most interesting cylindric billiards are particles with some hard ball pair-interactions, especially the system of Hard Balls. Though these systems possess some remarkable symmetries, their study gets more and more complicated as we increase the number of interacting balls. Ergodicity of two balls in any dimensions follows directly from the fundamental theorem as this system is a dispersive billiard ([2, 8]). Three balls are only semi-dispersive, thus their ergodicity is a much more difficult task ([9]). Generalization to the case of four balls has been achieved only with the dimensional restriction $\nu \geq 3$ on our balls ([10]). Later on even the ergodicity of an arbitrary number of hard balls has been proven ([13]), however the dimensional restriction $\nu \geq N$ was needed.

As for the hyperbolicity of Hard Ball Systems, applying highly nontrivial techniques in the geometric-algebraic part Simányi and Szász managed to discuss all possibilities for N and ν. Nevertheless the result in [18] only holds apart from a countable union of proper, analytic submanifolds for the outer geometric parameters (i.e. for the masses and radii of the particles).

Ergodicity and/or hyperbolicity *for some classes of hard ball systems* follows from the results of this paper. Any system of three interacting particles (with a connected collision graph) is a special case of the billiard discussed in section 3 (see also subsection 2.1). Thus Theorem 2.4 can be viewed in a way as a generalization of [9] to a cylindric billiard setting. A nice example of a Hamiltonian system with restricted hard ball interactions that belongs to the class of cylindric billiards discussed in section 4 is the one already mentioned in subsection 2.1, i.e. four balls with only one radius different from zero. A natural question is what we can say about the classical case of four hard balls (with all possible pairwise interactions allowed). For this system even ergodicity has been shown in [10] if $\nu \geq 3$, thus we focus our analysis on showing hyperbolicity for $\nu = 2$.

As to the dynamical-topological part, the analogue of Lemma 4.4 has been settled for any number of balls in any dimensions by a weak ball avoiding Lemma which uses strongly the symmetries of hard ball systems, together with the induction hypothesis (see e.g. section 5 from [18]). Thus in the rest of the appendix our aim is to demonstrate how the proof of Proposition 4.1 can be adapted to four hard balls. Our basic reference is section 4 from [10]. Rich collision sequences for the system of four balls were classified there (up to equivalence) in eleven cases. Cases 9-11 are analogous to cases (a) and (b) from Proposition 4.1 of this paper. Handling these sequences is much simplified by the fact that they contain subsequences sufficient with respect to a suitable sub-billiard dynamics. This is the strategy used in [10] as well; we may repeat the argumentation word by word with the only difference that because of $\nu = 2$ we get $codim(N) \geq 1$ instead of $codim(N) \geq 2$ for the submanifold of non-sufficient points.

In cases 1-8 however the proofs of [10] do not go through as they use the dimensional restriction (see also Remark 4.28. in that paper). These collision sequences are similar to our case (e) from Proposition 4.1 (actually case 1 is exactly our case (e)). Observe that for a cylindric billiard equivalent to a hard ball system it is always true that either $L_a \cap A_b = \{0\}$ or $L_a \perp L_b$, thus we need not care about situations analogous to subcase (e2) above. For brevity we have chosen one from cases 2-8, namely case 7, for which we show $codim(N) \geq 1$ by an adaptation of our Lemma 4.3. All the other collision sequences are treated in a much similar way.

In case 7 the collision sequence is $(\{3,4\}, \{1,4\}, \{1,2\}, \{1,3\}, \{1,4\})$ (here $\{i,j\}$ means a collision of the two balls $i \neq j$). In our notation we closely follow Lemma 4.3. Thus the advances are:
α' for the collision $\{3,4\}$;
α for the collision $\{1,3\}$;
β for the first collision $\{1,4\}$ and β' for the second.
Moreover we choose $\{1,2\}$ as our central collision with advance 0. The non-collision time moments we fix are:

t_* after $\{3,4\}$ and before the first $\{1,4\}$;
t_- after the first $\{1,4\}$ and before $\{1,2\}$;
t_+ after $\{1,2\}$ and before $\{1,3\}$;
$t_\#$ after $\{1,3\}$ and before the second $\{1,4\}$.
The form of a velocity vector v at any time moment is:

$$v = (v_1, v_2, v_3, v_4)$$

where, by the convention of zero total momentum we have $v_1 + v_2 + v_3 + v_4 = 0$ (here v_i means the two-dimensional velocity vector of the ball i). We calculate the neutral spaces at one of the time moments t_- and t_+ (by zero advance neutral vectors do not change at all in the course of the reflection $\{1,2\}$, thus the two neutral spaces coincide).

Neutrality with respect to the central collision $\{1,2\}$ means $w_1 = w_2$. Moreover, by the form of the generator subspace for any collision in a hard ball system and by the general formula (2.3), we get the following four equations for our neutral vector:

$$\begin{aligned}
w_3 - w_4 &= (\alpha' - \beta)(v_3^* - v_4^*) + \beta(v_3^- - v_4^-);\\
w_1 - w_4 &= \beta(v_1^- - v_4^-);\\
w_1 - w_3 &= \alpha(v_1^+ - v_3^+);\\
w_1 - w_4 &= (\beta' - \alpha)(v_1^\# - v_4^\#) + \alpha(v_1^+ - v_4^+).
\end{aligned} \tag{A.1}$$

By the second and the fourth of the above equations we know that for non-sufficient points (let us assume for a while $\beta \neq 0$):

$$\exists \lambda \in \mathbf{R}; \lambda \neq 0 : \quad \left((v_1^- - v_4^-) - \lambda(v_1^+ - v_4^+)\right) \parallel (v_1^\# - v_4^\#) \tag{A.2}$$

while if we subtract the second equation in (A.1) from the third, then together with the first equation we get that *for the same* λ:

$$\left((v_1^- - v_3^-) - \lambda(v_1^+ - v_3^+)\right) \parallel (v_3^* - v_4^*). \tag{A.3}$$

We may handle the degenerate possibilities of zero advances ($\alpha = 0$, $\beta = 0$) or parallelity (e.g. if $(v_1^- - v_4^-) \parallel (v_1^+ - v_4^+)$) exactly the same way as in Lemma 4.3. Otherwise apply a purely configurational translation

$$\delta q = (\delta q_1, \delta q_1, \delta q_1, -3\delta q_1)$$

at time t_-. By the structure of the generator spaces neither of the velocity vectors $v^-, v^+, v^\#$ changes, thus if the point remains nonsufficient, by (A.2), the value λ does not change throughout the perturbation (remember parallelity is excluded). As a consequence the left hand side of (A.3) remains constant, thus for the perturbed velocity $\bar{v}^*$ the difference $\bar{v}_3^* - \bar{v}_4^*$ is parallel to the original $v_3^* - v_4^*$. This component however moves on an arc of a circle.

Acknowledgements

First of all I am much greatful to my supervisor, Domokos Szász for suggesting me this topic and helping me a lot in all my considerations. Careful reading of the manuscript and lots of useful remarks are thankfully acknowledged for Nándor Simányi, Imre Péter Tóth and for both refrees of the paper.
This research was partially supported by Hungarian National Foundation for Scientific Research, grant OTKA T26176; and by the Research Group Stochastics@TUB of the Hungarian Academy of Sciences, affiliated to the Technical University of Budapest.

References

[1] D. Burago, S. Ferleger, A. Kononenko; *Uniform Estimates on the Number of Collisions in Semi-Dispersive Billiards*, Annals of Mathematics **147**, 695–708, (1998)

[2] Ya.G. Sinai and N. Chernov; *Ergodic Properties of Certain Systems of 2–D Discs and 3–D Balls*, Russian Math. Surveys **(3) 42**, 181–207, (1987)

[3] N. Chernov and C. Haskell; *Nonuniformly Hyperbolic K-systems are Bernoulli*, Ergod. Th. Dyn. Sys. **16**, 19–44, (1996)

[4] G. Galperin; *On Systems of Locally Interacting and Repelling Particles Moving in Space*, Trudy MMO **43**, 142-196, (1981).

[5] W. Hurewicz and H. Wallman; *Dimension Theory*, Princeton University Press, 1941

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

[7] A. Krámli, N. Simányi and D. Szász; *Ergodic Properties of Semi-Dispersing Billiards I. Two cylindric scatterers in the 3-D torus*, Nonlinearity **2**, 311–326, (1989).

[8] A. Krámli, N. Simányi and D. Szász; *A "Transversal" Fundamental Theorem for Semi-Dispersive Billiards*, Comm. Math. Phys. **129**, 535–560, (1990).

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

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

[11] C. Liverani and M. Wojtkowski; *Ergodicity in Hamiltonian Systems*, Dynamics Reported **4** (New series), 130–202, (1995).

[12] D. Ornstein and B. Weiss; *On the Bernoulli nature of systems with some hyperbolic structure*, Ergod. Th. Dyn. Sys. **18**, 441-456, (1998).

[13] N. Simányi; *The K-Property of N Billiards Balls I.*, Invent. Math. **108**, 521–548, (1992); *II.* ibidem **110**, 151–172, (1992).

[14] N. Simányi; in preparation

[15] N. Simányi, Oral communication

[16] N. Simányi and D. Szász; *The K-Property of* $4-D$ *Billiards with Non-Orthogonal Cylindric Scatterers*, J. Stat. Phys. **76, Nos. 1/2**, 587–604, (1994).

[17] N. Simányi and D. Szász; *The K-Property of Hamiltonian Systems with Restricted Hard Ball Interactions*, Mathematical Research Letters **2, No. 6**, 751–770, (1995).

[18] N. Simányi and D. Szász; *Hard Ball Systems Are Completely Hyperbolic*, Annals of Mathematics **149**, 35–96, (1999).

[19] N. Simányi and D. Szász; *Non-Integrability of Cylindric Billiards and Transitive Lie-Group Actions*, to appear in Ergod. Th. Dyn. Sys.

[20] D. Szász; *Ergodicity of Classical Billiard Balls*, Physica A **194**, 86–92, (1993).

[21] D. Szász; *The K-Property of "Orthogonal" Cylindric Billiards*, Comm. Math. Phys. **160**, 581–597, (1994).

[22] D. Szász; *Boltzmann's Ergodic Hypothesis, a Conjecture for Centuries?*, Studia Sci. Math. Hung **31**, 299–322, (1996).

[23] D. Szász; *Ball-Avoiding Theorems*, to appear in Ergod. Th. Dyn. Sys.

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