Journal of Statistical Physics, Vol. 76, Nos. 1/2, 1994

The K-Property of 4D Billiards with Nonorthogonal Cylindric Scatterers

Nándor Simányi[1] **and Domokos Szász**[1]

Received September 30, 1993; final January 27, 1994

The K-property of cylindric billiards given on the 4-torus is established. These billiards are neither "orthogonal," where general necessary and sufficient conditions were obtained by D. Szász, nor isomorphic to hard-ball systems, where the connecting path formula of N. Simányi is at hand.

KEY WORDS: Local and global ergodicity; hyperbolic dynamical systems; semidispersing billiards; cylindric billiards; hard-ball systems.

1. INTRODUCTION

Motivated by a question of John Mather, a systematic study of *toric billiards with cylindric scatterers* was initiated in ref. 11. Since this class of billiards is relatively simple, one can hope for general necessary and sufficient conditions for the ergodicity (and the K-property) of these systems. Indeed, in ref. 12 necessary and sufficient conditions were obtained for the ergodicity of a subclass: the family of *orthogonal* cylindric billiards.

Despite the simplicity of the category, to obtain a complete answer does not promise to be very easy because cylindric billiards contain hard-ball systems as well. For them, however, the celebrated Boltzmann–Sinai ergodic hypothesis is still not settled in general. For the reader interested in the relation of cylindric billiards and hard-ball systems on one hand, and in the recent status of the Boltzmann–Sinai ergodic hypothesis on the other hand, we will return to these questions after the formulation of the main result of the present work.

In this paper we consider cylindric billiards on the 4-torus with two nonorthogonal cylindric scatterers. For brevity of exposition, we only treat

[1] This paper is dedicated to Philippe Choquard on the occasion of his 65th birthday. Mathematical Institute of the Hungarian Academy of Sciences, H-1364, Budapest, Hungary.

0022-4715/94/0700-0587$07.00/0

cases when the bases of these cylinders are 2D discs. The philosophy behind this restriction is that, in general, the higher the dimensions of the bases are, the easier the proof of ergodicity becomes.

Consider on a 4-torus $\mathbb{T}^4 := \mathbb{R}^4/D$ two cylinders $C^j := \{x \in \mathbb{T}^4 : \operatorname{dist}(x, A^j) \leqslant r^j\}$, $j = 1, 2$. Here D is an arbitrary rank-four discrete subgroup (a lattice) in the Euclidean space $\mathbb{R}^4$ and the metric "dist" is inherited from the usual Euclidean metric of $\mathbb{R}^4$ by factorization with respect to D. The "axes" A^j of these cylinders C^j are supposed to be cosets with respect to two transversal 2-dimensional subtori of $\mathbb{T}^4$. Without loss of generality we can assume that these tori A^j contain the origin, i.e., they themselves are two transversal 2-dimensional subtori.

Denote $\mathbf{Q} := \mathbb{T}^4 \backslash (C^1 \cup C^2)$ and $M := \mathbf{Q} \times \mathbf{S}^3$, where $\mathbf{S}^d$ is the unit d-sphere. $M = \{x = (Q, V)\}$ is the phase space of the billiard given in the domain $\mathbf{Q}$ possessing cylindric boundaries. The dynamical system $(M, S^{\mathbb{R}}, d\mu)$, where $S^{\mathbb{R}}$ is the dynamics defined by uniform motion inside the domain and specular reflections at its boundary (the scatterers!) and $d\mu$ is the Liouville measure, is the *cylindric billiard* we want to study. (As to notions and notations in connection with semidispersing billiards we follow ref. 4.) To avoid unnecessary complications we assume that the radii r^1, r^2 are sufficiently small, so that the configuration space is connected.

The cardinal example of a billiard with nonorthogonal cylindric scatterers—fulfilling all conditions of the upcoming theorem—is the billiard system of three disks in the two-dimensional torus $\mathbb{T}^2$, where only the first disk interacts (collides elastically) with the others, while the interaction of the other two disks with each other is "canceled" and they are allowed to overlap.

Our main result is the following:

Theorem. A sufficient condition [besides the already mentioned transversality $\dim(A^1 \cap A^2) = 0$] for the K-property of the cylindric billiard introduced before is that A^1 be transversal to the orthogonal complement L^2 of A^2 or, equivalently, A^2 be transversal to the orthogonal complement L^1 of A^1. (Here A^j is identified with the corresponding subspace of $\mathbb{R}^4$.)

As indicated before, we—for the interested reader—are going to present some remarks on the relation of cylindric billiards to hard-ball systems and provide a summary of earlier results.

Assume that, in general, a system of N ($\geqslant 2$) balls of radii $r > 0$ are given on $\mathbf{T}^\nu$, the ν-dimensional unit torus ($\nu \geqslant 2$). Denote the phase point of the ith ball by $(q_i, v_i) \in \mathbf{T}^\nu \times \mathbf{R}^\nu$. The configuration space $\tilde{\mathbf{Q}}$ of the N balls is a subset of $\mathbf{T}^{N \cdot \nu}$: from $\mathbf{T}^{N \cdot \nu}$ we cut out $\binom{N}{2}$ cylindric scatterers:

$$\tilde{C}_{i,j} = \{Q = (q_1, \ldots, q_N) \in \mathbf{T}^{N \cdot \nu} : |q_i - q_j| < 2r\}$$

$1 \leqslant i < j \leqslant N$. The energy $H = \frac{1}{2}\sum_1^N v_i^2$ and the total momentum $P = \sum_1^N v_i$ are first integrals of motion. Thus, without loss of generality, we can assume that $H = \frac{1}{2}$ and $P = 0$ and, moreover, that the sum of spatial components $B = \sum_1^N q_i = 0$ (if $P \neq 0$, then the center of mass has an additional conditionally periodic or periodic motion). For these values of H, P, and B, the phase space of the system reduces to $M := \mathbf{Q} \times \mathscr{S}^{N \cdot \nu - \nu - 1}$, where

$$\mathbf{Q} := \left\{ Q \in \tilde{\mathbf{Q}} : \sum_1^N q_i = 0 \right\}$$

with $d := \dim \mathbf{Q} = N \cdot \nu - \nu$, and $\mathscr{S}^k$ denotes, in general, the k-dimensional unit sphere. It is easy to see that the dynamics of the N balls, determined by their uniform motion with elastic collisions on one hand, and the billiard flow $\{S^t : t \in \mathbf{R}\}$ in $\mathbf{Q}$ with specular reflections at $\partial\mathbf{Q}$ on the other hand, conserve the Liouville measure $d\mu = \text{const} \cdot dq \cdot dv$ and are isomorphic.

In the aforementioned billiards the smooth components $\partial\mathbf{Q}_i$ of the boundary of $\mathbf{Q}$ are convex from inside the domain. We can, in general, consider billiards in bounded, connected, closed domains on the d-torus $\mathbf{T}^d$. We say that such a billiard is *semidispersing* (*dispersing*) if the smooth components of the boundary are *convex* (*strictly convex*) from inside the domain. Thus the system of two hard balls is dispersing while that of $N \geqslant 3$ balls is semidispersing. The dispersing (semidispersing) property of these billiard lies behind the proofs of their ergodicity (and in all cases automatically of their K-property as well).

A brief summary of the results:

1. In 1970, Sinai, in his celebrated work,[7] proved the K-property of two two-dimensional discs ($N = 2$, $\nu = 2$; see also ref. 2).
2. In 1987, Chernov and Sinai[10] established the K-property of two balls on an arbitrary-dimensional torus ($N = 2$, $\nu > 2$).
3. In 1989, Krámli *et al.*[3] considered a three-dimensional orthogonal cylindric billiard; they obtained its K-property and thus this was the first semidispersing—but not dispersing—billiard whose ergodicity was shown.
4. In 1990, Krámli *et al.*[4] demonstrated the K-property of $N = 3$ balls on the ν-torus whenever $\nu \geqslant 2$.
5. In 1991, Krámli *et al.*[5] improved their methods to get the ergodicity of $N = 4$ balls on the ν-torus whenever $\nu \geqslant 3$.
6. In 1992, Bunimovich *et al.*[7] introduced a Hamiltonian model of an arbitrary number of balls in arbitrary-dimensional space; the walls of the domain where these balls can move are convex

from inside the domain and the authors were able to exploit the resulting hyperbolicity to derive the K-property. Their proof also needed an additional constraint: the invariance of the order of balls. As to this invariance, the model is similar to a toy model of Chernov and Sinai, the so-called *pencase*.

7. In 1992, Simányi[9] was able to establish the strongest result yet for hard-ball systems: the system of $N \geqslant 2$ balls is ergodic on the ν-torus whenever $\nu \geqslant N$.
8. In 1993, Szász[11, 12] started a systematic study of cylindric billiards and found a sufficient and necessary condition for the ergodicity of a class of them: the orthogonal cylindric billiards.

Return to the model of the present paper. The proof of our theorem is based on the strategy formulated in refs. 3 and 5. Indeed, in the spirit of the latter work, the proof of global ergodicity of a semidispersing billiard should be based on a suitable definition of richness and then essentially consist of three parts:

1. Geometric-algebra part for treating neighborhoods of rich points.
2. Dynamical-topological part for handling the subset of nonrich points.
3. Finally, separate arguments for singular trajectories (also settling the Chernov–Sinai Ansatz).

Though our definition of richness, in its spirit, is analogous to the one used in ref. 12, the arguments of that work are not applicable here. There the method is based on the abundance of pairs of different cylinders such that their base spaces (like L^1 and L^2 in our context) have nontrivial intersections and this property was used throughout the whole proof. Closer to our situation is the case of $N=3$ balls on the 2-torus[4] where the configuration space of the isomorphic billiard is four-dimensional and there are three cylindric scatterers each with a two-dimensional base. We can, indeed, use several ideas of that work, though still there is a fundamental difference: the scatterers of the hard-ball billiard inherit the permutation invariance of the balls. This invariance led, in particular, in ref. 9 to the connecting path formula that, at least according to our knowledge, is missing for general cylindric billiards.

The paper is organized as follows. In Section 2 we present the notion of richness, formulate the main lemmas, and indicate how they imply the theorem. Section 3 is devoted to the geometric-algebraic part 1 by proving Main Lemma 2.2. Section 4 then settles the dynamical-topological part 2 by proving Main Lemma 2.3 and also contains the necessary remarks as to part 3.

For brevity, our exposition relies heavily on earlier work.[3–6, 12] To help the reader, however, we everywhere provide precise references to the occurrence of the necessary definitions, statements, and arguments.

2. NOTION OF RICHNESS, MAIN LEMMAS, AND THE PROOF

Similarly to ref. 5, M^* will denote the set of phase points whose orbits contain an infinite number of collisions among which not more than one is singular. $M^0 \subset M^*$ will be the subset of regular phase points, and $M^1 := M^* \backslash M^0$. Moreover, $\mathscr{SR}^+$ will denote the collection of all phase points $x \in \partial M$ for which the reflection, occurring at x, is singular (tangential or multiple) and, in the case of a multiple collision, x is supplied with the *outgoing* velocity V^+. We remind the reader that a trajectory segment $S^{[a,b]}x$ is called *regular* (or nonsingular) if it does not hit singularities ($S^{[a,b]}x \cap \mathscr{SR}^+ = \emptyset$; cf. ref. 4).

Consider the regular trajectory segment $S^{[a,b]}x$, $-\infty \leqslant a < b \leqslant \infty$, $x \in M$. Its symbolic collision sequence is the list of subsequent cylinders of collisions $(C^{j_1},..., C^{j_k})$, $1 \leqslant k$, of the trajectory and can be described by the sequence $(j_1,..., j_k)$, $j_l = 1, 2$, $1 \leqslant l \leqslant k$. (If the trajectory hits one or more singularities, then, of course, there are a finite number of such sequences for any finite orbit.) The reader is reminded of the concept of island from ref. 5: it is a maximal subsequence of the symbolic collision sequence consisting of consecutive collisions with the same cylinder.

Definition 2.1. We say that the trajectory segment $S^{[a,b]}x$ is *rich* if its symbolic collision sequence contains at least four islands. If the trajectory segment hits singularities, then this property is required for any trajectory branch.

Finally, the trajectory segment is *poor* if it is not rich.

Main Lemma 2.2. Assume that the trajectory segment $S^{[a,b]}x$ is regular, $S^a x$, $S^b x \notin \partial M$, and its trajectory segment is rich. Then there exist a neighborhood $U \subset M$ of x, and a submanifold N such that:

1. codim $N \geqslant 2$.
2. For every $y \in U \backslash N$, $S^{[a,b]}y$ is sufficient.

(As to the definition of sufficiency, see Definition 2.4 of ref. 6.)

The demonstration of Main Lemma 2.2 will be the content of Section 3.

Denote by M_p^0 the subset of poor phase points from M^0. (A phase point is called poor if its entire trajectory is poor.) There may exist some trivial codimension-one submanifolds of nonsufficient points for our billiard.

Therefore we should exclude a finite union of codimension-one submanifolds to obtain $M^{\#}$.

Consider the subbilliard $(M^j, (S^j)^{\mathbb{R}}, d\mu^j)$, $j = 1, 2$, on the 4-torus with the cylinder C^j as the only scatterer. By Lemma A.2.1 from ref. 12, for fixed $j = 1, 2$, the subset $\tilde{M}^j$ of phase points from M^j whose orbits have no collisions at all is contained in a finite union $\tilde{M}^j := \bigcup_{l=1}^{l^j} E_l$ of submanifolds such that for every l: $1 \leqslant l \leqslant l^j$:

1. We have

$$\operatorname{codim}_{M^j} E_l \geqslant 1$$

2. E_l^{cl} is compact.

Since M, M^j $(j = 1, 2) \subset \mathbf{T}^4 \times S^3$, it makes sense to consider the connected components of the set

$$M^{\#} := M \setminus (\tilde{M}^1 \cup \tilde{M}^2)$$

By Lemma A.2.1 of ref. 12 their number is finite. Denote them by $\Omega_1,..., \Omega_I$ $(1 \leqslant I < \infty)$.

Now we can claim:

Main Lemma 2.3. $M_p^0 \cap M^{\#} \cap M_{ns}$ is a residual subset of M.

The demonstration can be found in Section 4. Recall the meaning of the phrase "residual set": Such a set can be covered by a countable collection of codimension-two $(\geqslant 2)$ closed sets with zero measure. M_{ns} denotes the set of all nonsufficient (nonhyperbolic) phase points.

The reader familiar with the technique of establishing global ergodicity of semidispersing billiards already knows that the treatment corresponding to the previous main lemmas for singular points on one hand, and the verification of the Chernov–Sinai Ansatz on the other hand, follow from the next result:

Main Lemma 2.4. For every cell C of maximal dimension 5 in $\mathscr{SR}^+$, the set $C_{es} \subset C$ of all eventually simple phase points can be covered by a countable family of closed zero-subsets (with respect to the surface measure μ_C in C) of C.

We say that a point $x \in \mathscr{SR}^+$ is *eventually simple* if:

1. The semitrajectory $S^{\mathbb{R}_+}x$ is regular.
2. There is a number $t_0 > 0$ such that the trajectory segment $S^{(t_0, \infty)}x$ contains not more than one island.

As said earlier, Main Lemma 2.4 can be derived in almost exactly the same way as Main Theorem 6.1 in ref. 6, and, for brevity, we are satisfied by giving some indications at the end of Section 4.

Given Main Lemmas 2.2–2.4, the proof of our theorem is not difficult. First of all, the transversality of A^1 and A^2 is necessary. For if $\dim(A^1 \cap A^2) > 0$, then by ref. 3 and by ref. 7, respectively, the system is the product of a K-system and a conditionally periodic motion and, consequently, is not K.

Assume thus that $\dim A^1 \cap A^2 = 0$ and A^1 is transversal to the orthogonal complement L^2 of A^2, which is just to require that the subspaces A^1 and A^2 are in duality with respect to the Euclidean scalar product of $\mathbb{R}^4$. It is an easy consequence of these conditions that the orthocomplements L^1 and L^2 are transversal as well, and they are also in duality with respect to the Euclidean scalar product.

Since our main lemmas were formulated completely identically to those in ref. 12, the way the theorem follows from them is also the same and will be omitted. We only note that the traditional strategy also applied in ref. 12 first implies that each component $\Omega_1,..., \Omega_l$ belongs to one ergodic component. Then the completion of the proof, i.e., showing that these components form, in fact, exactly one ergodic component, follows from Lemma A.2.3 of ref. 12 as shown in the paper just mentioned. Finally, we note that Sinai's transversality condition holds true for the intersection of the surfaces of the cylinders C^j: the normal vectors to these surfaces are always nonparallel, for they belong to the transversal subspaces L^j. By Sinai's classical result[8] this ensures that the moments of collisions cannot accumulate in bounded time intervals. ■

3. GEOMETRIC-ALGEBRAIC CONSIDERATIONS: PROOF OF MAIN LEMMA 2.2

Throughout the whole section we will only consider regular trajectory segments.

Proof of Main Lemma 2.2. Consider first a bounded trajectory segment $S^{[a,b]}x_0$ ($a<0$, $b>0$) with the island structure $(1, 2)$ such that (for simplicity) $t=0$ separates the collisions with C^1 and C^2. Consider, moreover, an arbitrary configuration-displacement vector ΔQ neutral with respect to the segment $S^{[a,b]}x_0$. The corresponding time shifts (advances) are denoted by α and β. (For their definition see, for example, ref. 9, I, Section 2.) Here we need to introduce four linear operators of $\mathbb{R}^4$: P_j is the orthogonal projection onto the subspace L^j ($j=1, 2$), while π_j is the (nonorthogonal) projection of $\mathbb{R}^4$ onto A^j corresponding to the direct sum decomposition $\mathbb{R}^4 = A^1 + A^2$. The neutrality of ΔQ with respect to island 1 (with advance α) precisely means that $P_1(\Delta Q - \alpha V_0) = 0$, that is, $\Delta Q - \alpha V_0 = a_1 \in A^1$. (Here V_0 denotes the velocity vector of x_0.) Similarly, the neutrality with respect to island 2 with advance β means that $\Delta Q - \beta V_0 = a_2 \in A^2$. By subtraction $(\alpha - \beta) V_0 = a_2 - a_1$, so $a_1 = (\beta - \alpha)\pi_1(V_0)$, $a_2 = (\alpha - \beta)\pi_2(V_0)$. For ΔQ we obtain the following symmetric pair of formulas:

$$\Delta Q = \alpha V_0 + (\beta - \alpha)\pi_1(V_0) = \beta V_0 + (\alpha - \beta)\pi_2(V_0) \tag{3.1}$$

This formula immediately shows that *the dimension of the space of neutral vectors (with respect to $S^{[a,b]}x_0$) is two.* (Together with the flow direction.)

Let us now suppose that there is an additional island of type 1 following the above considered two, and the separating moment of time between the last two islands is t_1. We keep all notations of the arguments resulting in (3.1). The advance of the new island is denoted by γ. From (3.1) and the basic properties of the flow we get that the investigated configuration displacement ΔQ *at time* t_1 is

$$\Delta Q(t_1) = \beta V_{t_1} + (\alpha - \beta)\pi_2(V_0) \tag{3.2}$$

On the other hand, by the analog of (3.1) applied to the second pair of islands, we have

$$\Delta Q(t_1) = \beta V_{t_1} + (\gamma - \beta)\pi_2(V_{t_1}) \tag{3.3}$$

Now by subtraction

$$(\alpha - \beta) V_0 + (\beta - \gamma) V_{t_1} \in A^1 \tag{3.4}$$

The advantage of the last formula is that it forces the three advances to be equal (and, therefore, gives the required sufficiency of the trajectory segment!) whenever the velocity vectors V_0 and V_{t_1} are *linearly independent over the subspace* A^1, that is, they are independent and they span a subspace transversal to A^1. If this is not the case, then the velocity change $V_{t_1} - V_0$ is in the subspace spanned by V_0 and A^1, denoted by span (V_0, A^1). But, by the nature of the dynamics involved, this difference must also belong to the subspace L^2. However, the assumed tranversality between A^1 and L^2 implies that the intersection

$$\text{span}(V_0, A^1) \cap L^2 \text{ is a line} \tag{3.5}$$

This line is determined by the billiard table and V_0. Let us now apply a pure configuration translation to the original phase point x_0 such that the vector describing this translation is in L^2 but it is not parallel with $P_2(V_0)$. Under the effect of such translations the velocity difference $V_{t_1} - V_0$ moves along an arc of a circle in L^2. This observation, together with (3.5), shows that *outside of some codimension-one smooth submanifold of a neighborhood of* x_0 *the trajectory segment island structure* (1, 2, 1) *is sufficient.*

Let us go further and suppose that a brand new island of type 2 precedes the collision sequence (1, 2, 1) studied above. Applying the above arguments, we get that the condition for the nonsufficiency of the "left part" (2, 1, 2) is the following one:

$$V_{t_2} - V_0 \in \operatorname{span}(V_0, A^2) \cap L^1 \tag{3.6}$$

where

$$\operatorname{span}(V_0, A^2) \cap L^1 \text{ is a line} \tag{3.7}$$

(Here V_{t_2} is the velocity at some moment t_2 between the first two islands.) Now (3.6) determines a codimension-one smooth submanifold Σ^-, just like the former relationship:

$$V_{t_1} - V_0 \in \operatorname{span}(V_0, A^1) \cap L^2 \tag{3.8}$$

which defines the codimension-one submanifold Σ^+. Our task is now to show that Σ^- and Σ^+ *are transversal submanifolds.*

We will easily find a pure configuration displacement tangent vector ΔQ of M, at $x_0 \in \Sigma^- \cap \Sigma^+$ that is tangent to Σ^+ but is not tangent to Σ^-. Obviously, the first requirement is fulfilled by any $\Delta Q \in A^2$, while the second one will be valid whenever $\Delta Q \notin \operatorname{span}(V_0, A^1)$. By the assumed transversality between A^1 and A^2, the intersection of the subspaces A^2 and $\operatorname{span}(V_0, A^1)$ is one-dimensional; therefore it is certainly possible to choose a vector ΔQ with the required properties.

The proof of Main Lemma 2.2 is now complete. ∎

4. DYNAMICAL-TOPOLOGICAL PART: PROOF OF MAIN LEMMA 2.3

From our definition of richness it follows that poor phase points undergo not more than three collisions during their whole trajectory; thus $M_p^0 \cap M^\# := M_1^0 \cap M_2^0 \cap M_3^0$, where M_i^0 denotes the set of regular trajectories containing exactly i islands ($i = 0, 1, 2,\ldots$). (Of course, $M_0^0 \cap M^\# = \varnothing$.)

The residuality of M_1^0 follows by a trivial adaptation of the proof of sublemma 2 of ref. 3. Thus Main Lemma 2.3 will immediately come from our next two lemmas.

Lemma 4.1. M_2^0 is a residual subset.

Lemma 4.2. $M_3^0 \cap M_{ns}$ is a residual subset.

The proofs of these lemmas are analogous to the proof of Main Lemma 5.1 from ref. 5 and we only give the necessary vocabulary and supplements to adapt it to our situation.

Proof of Lemma 4.1. It is sufficient to show that any $x^0 = (Q^0, V^0)$ ($x^0 \notin \partial M$) has a neighborhood U such that $U \cap M_2^0$ is residual. We can and do assume that $x^0 \in M_2^0 \backslash \partial M$. For simplicity we suppose that $S^{\mathbb{R}^-} x^0$ has just one island of type C^1 and $S^{\mathbb{R}^+} x^0$ has just one island of type C^2. Choose a small open ball neighborhood $U \subset M \backslash \partial M$ of x^0. We denote

$$F_- := \{x \in U: (S^{\mathbb{R}^-} x) \cap \operatorname{int}(\partial M^2) = \varnothing\} \tag{4.1}$$

$$F_+ := \{x \in U: (S^{\mathbb{R}^+} x) \cap \operatorname{int}(\partial M^1) = \varnothing\} \tag{4.2}$$

[Here $\operatorname{int}(\partial M^j)$ denotes the part of the boundary ∂M of M that corresponds to the *proper* (nontangential) collisions with the cylinder C^j.]

First we construct the invariant manifolds, the fundamental tools of the proof. Of course, for any $x = (Q, V) \in U$

$$\gamma_0^s(x) := \{(Q', V) \in U : Q' - Q \in A^2\},$$

$$\gamma_0^u(x) := \{(Q', V) \in U : Q' - Q \in A^1\}$$

In constructing $\gamma_{\exp}^u(x)$, instead of using L_{q_2, v_2} of ref. 5, here we need the three-dimensional submanifold $L(Q - P_1(Q), V - P_1(V))$ that can be obtained by fixing the orthogonal projections of Q and V onto the constituent space A^1 of C^1. [The operator $I - P_1$ is just the projection onto A^1. It is appropriate to admit here that, being in a torus instead of a Euclidean space, the configuration projection $Q - P_1(Q)$ is only defined locally, but this is just fine for our local analysis in U.] Then

$$\gamma_{\exp}^u(x) := \{y \in L(Q - P_1(Q), V - P_1(V)) \cap U: \operatorname{dist}\{(S^1)^t x, (S^1)^t y\} \to 0$$
$$\text{exponentially fast as } t \to -\infty\}$$

Analogously, in the definition of $\gamma_{\exp}^s(x)$, we use the submanifold $L(Q - P_2(Q), V - P_2(V))$. Thus

$$\gamma_{\exp}^s(x) := \{y \in L(Q - P_2(Q), V - P_2(V)) \cap U: \operatorname{dist}\{(S^2)^t x, (S^2)^t y\} \to 0$$
$$\text{exponentially fast as } t \to \infty\}$$

The pseudo-(un)stable invariant manifolds are defined by (5.6) of ref. 5, and

$$F'_{+} := \{x \in U: (S^{\mathbb{R}^+}x) \cap \operatorname{int}(\partial M^1) = \emptyset \text{ if the radius } r^1 \text{ is shrunk to } r^1 - \varepsilon\} \tag{4.3}$$

$$F'_{-} := \{x \in U: (S^{\mathbb{R}^-}x) \cap \operatorname{int}(\partial M^2) = \emptyset \text{ if the radius } r^2 \text{ is shrunk to } r^2 - \varepsilon\} \tag{4.4}$$

Now the analogs of Lemmas 5.3, 5.8, and 5.9 of ref. 5 are automatically valid. Moreover, $\dim \gamma^{s,u}_{\exp} = 1$, $\dim \gamma^0 = 4$, and one dimension is missing here, too. The solution is again similar to ref. 5. For the proof of the residuality of $F_+ \cap F_-$ we need to prove the following proposition:

Proposition 4.3. For almost every point $x \in U$, almost every other point y in a suitably small neighborhood of x can be connected with x by a sequence of sequentially intersecting (i.e., the neighboring members are intersecting) pseudostable and pseudounstable manifolds.

Here the key issue is the missing seventh dimension from the span of the tangent spaces of these pseudoinvariant manifolds. The problem only arises in connection with the velocity variations, for γ^0 alone already contains all pure spatial variations. So let us focus on what happens to the velocity vector $V \in S^3$ while executing exponential unstable and stable perturbations corresponding to $\gamma^u_{\exp}$ and $\gamma^s_{\exp}$. The effect of the first one on V is just some *rotation* around the subspace A^1, while the second type of variation just causes rotation around the subspace A^2. The whole problem will be settled and Main Lemma 4.1 will be proved if we prove the following geometric lemma concerning some special transformation groups:

Lemma 4.4. Using the notations of this paper, assume that A^1 is transversal to both A^2 and its orthogonal complement L^2. Let G_j be the group of all rotations around A^j, and denote by G the transformation group generated by G_1 and G_2. [It is known from the classical theory of Lie groups that the group G is a connected—not necessarily closed—Lie subgroup of $SO(4)$.] Then, the natural action of G on S^3 is transitive.

Remark. It is very appealing to conjecture that, under the conditions of this lemma, the group G is necessarily the whole special orthogonal group $SO(4)$. However, right after the proof of Lemma 4.4 we present a family of examples—involving the one mentioned right before the theorem in the introduction—showing that sometimes G is merely the four-dimensional Lie group $U(2)$.

Proof. First of all, we claim the following:

Sublemma 4.5. Under the conditions of Lemma 4.4 the G-action on S^3 has an open orbit with full measure. (Or, equivalently, it has an open and dense orbit.)

Proof of the Sublemma. Observe that the question about the openness of orbits is, by the very nature of things, a local one; therefore it is enough to consider the infinitesimal action of the Lie group G on the three-sphere. This means that we multiply (from the left) the points V of S^3 by the Lie algebra of G, and thus we obtain the tangent space of the orbit of V. In order to understand (to some extent) the Lie algebra of G, we consider an infinitesimal generator X_j of the one-parameter subgroup G_j. A possible choice for X_j is a rotation by the right angle (in one of the two possible directions) around A^j. Next we define a suitable base $\{e_1, f_1, e_2, f_2\}$ in $\mathbb{R}^4$ situated pretty symmetrically with respect to the pair of subspaces (L^1, L^2), so that the matrices of the generators X_j become rather simple in this base. Namely, set

$$a := \max\{\|P_2(x)\|: x \in L^1 \text{ and } \|x\| = 1\} \tag{4.5}$$

$$b := \min\{\|P_2(x)\|: x \in L^1 \text{ and } \|x\| = 1\} \tag{4.6}$$

By the assumptions of Lemma 4.4 we have $0 < b \leqslant a < 1$. We note that the relative position of the two-dimensional subspaces A^1 and A^2 (or, equivalently, the relative position of their orthocomplements L^1 and L^2) is completely characterized by the acute angles $\alpha_0 = \arccos(a)$ and $\beta_0 = \arccos(b)$. The first of these angles is the minimal possible angle between a nonzero vector of L^1 and the subspace L^2, while the second one is the maximum of such angles. Of course, the role of the two subspaces can be interchanged in these definitions. The vector e_1 is now chosen to be one of the two (antipodal) unit vectors of L^1 for which the maximum of (4.5) is attained. (In the case of $a = b$ any unit vector of L^1 would do.) The vector $e_2 \in L^2$ is the uniquely determined unit vector that makes an angle of α_0 with e_1. Note that e_2 has the same direction as the orthogonal projection $P_2(e_1)$ of e_1 onto L^2. The pair of unit vectors $f_j \in L^j$ is defined analogously: f_1 is one of the two opposite unit vectors of L^1 that makes the maximum angle β_0 with L^2 and f_2 is the uniquely determined unit vector of L^2 that makes the angle β_0 with f_1. By elementary properties of the quadratic forms, the pair $\{e_j, f_j\}$ is an orthonormal base in L^j and $\langle e_i, f_j \rangle = 0$ for $i, j = 1, 2$. It is now an easy task to write down the matrices of the rotations X_j in the base $\{e_1, f_1, e_2, f_2\}$:

$$X_1 = \begin{pmatrix} 0 & -1 & 0 & -b \\ 1 & 0 & a & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \qquad X_2 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & -b & 0 & -1 \\ a & 0 & 1 & 0 \end{pmatrix}$$

Of course, the Lie algebra of the group G is the Lie subalgebra of so(4) generated by X_1 and X_2. So next we compute the Lie bracket of these matrices:

$$[X_1, X_2] = \begin{pmatrix} -ab & 0 & -b & 0 \\ 0 & -ab & 0 & -a \\ b & 0 & ab & 0 \\ 0 & a & 0 & ab \end{pmatrix}$$

The products of the matrices X_1, X_2, $[X_1, X_2]$ with an arbitrary unit velocity vector

$$V = v_1 e_1 + w_1 f_1 + v_2 e_2 + w_2 f_2 \in S^3$$

are as follows:

$$X_1 V = \begin{pmatrix} -w_1 - bw_2 \\ v_1 + av_2 \\ 0 \\ 0 \end{pmatrix}, \qquad X_2 V = \begin{pmatrix} 0 \\ 0 \\ -bw_1 - w_2 \\ av_1 + v_2 \end{pmatrix},$$
$$[X_1, X_2]\, V = \begin{pmatrix} -b(av_1 + v_2) \\ -a(bw_1 + w_2) \\ b(v_1 + av_2) \\ a(w_1 + bw_2) \end{pmatrix} \tag{4.7}$$

If for some $V \in S^3$ the above three vectors are linearly independent, then the orbit of V is obviously open. The unwanted linear connectedness of the vectors in (4.7) is easy to check. If the point $V \in S^3$ does not belong to either of the subspaces A^j ($j = 1, 2$), then $X_1 V$ and $X_2 V$ are independent. However, the intersections $A^j \cap S^3$ are circles with zero measure, so they can be discarded. Once $X_1 V$ and $X_2 V$ are independent, the linear dependence of the third vector $[X_1, X_2]\, V$ on the first two can be checked separately on the first two and last two coordinates. In both cases the dependence means that a certain two-by-two determinant vanishes, and the resulting two equations *are the same*:

$$Q(V) := b(a(v_1^2 + v_2^2) + (a^2+1)\, v_1 v_2) + a(b(w_1^2 + w_2^2) + (b^2+1)\, w_1 w_2) = 0 \tag{4.8}$$

It is easy to see that the signature of the quadratic form Q is $(2, 2)$, and $\mathbb{R}^4$ is the Q-orthogonal (and Euclidean orthogonal, too) direct sum of the subspaces $H = \operatorname{span}(e_1, e_2)$ and $H^\perp = \operatorname{span}(f_1, f_2)$, and Q is indefinite in both subspaces. (The relations $0 < b$ and $a < 1$ are used here.) Due to the signature $(2, 2)$, the hypersurface $Q(V) = 0$ dissects the sphere S^3 into two *connected* open sets S^3_- and S^3_+ according to the sign of Q. We have proved that all points of the open (in S^3) set

$$\mathcal{O} = S^3 \setminus (\{V \in S^3 : Q(V) = 0\} \cup A^1 \cup A^2) \tag{4.9}$$

have open G-orbits. We observe here that it is not necessary to drop the circles $A^j \cap S^3$ $(\subset S^3_+)$ from S^3 in the formula (4.9), since for the points $V \in A^1 \cap S^3$ the vector $X_2 V$ is not tangential to the circle $A^1 \cap S^3$, so the action of the group G_2 moves this point V out of the mentioned circle, and a similar argument applies to the points of $A^2 \cap S^3$. Thus, so far we have seen that the open set $S^3_- \cup S^3_+$ is covered by the open orbits of the G-action on S^3. By the connectedness of $S^3_\pm$, each of these two sets belongs to one orbit. The last thing we need to do in order to prove Sublemma 4.5 is to connect S^3_- and S^3_+ by the G-action. The one-parameter subgroup of G generated by $[X_1, X_2]$ will give use the needed help. Namely, direct inspection shows that the subspace $H = \operatorname{span}(e_1, e_2)$ is invariant under the multiplication by $[X_1, X_2]$, and the restriction of this Lie bracket to H is nonzero. Therefore the action of the one-parameter subgroup generated by $[X_1, X_2]$ on H can only be the full special orthogonal group of H. Since the quadratic form is indefinite on H, the arising rotations of H transport points from S^3_- to S^3_+. Hence the sublemma.

Finishing the proof of Lemma 4.4. By Sublemma 4.5 there is an open and dense orbit of G. Since the orbits of the action of $\bar{G}$ (the closure of G) are the closures of the G-orbits, we have that $\bar{G}$ acts transitively on S^3. Therefore, every G-orbit is dense, so—by Sublemma 4.5 again—the action of G on S^3 is also transitive. Hence Lemma 4.4. ∎

Thus Lemma 4.1 is also established. ∎

Example. Make the usual identification between our real Euclidean space $\mathbb{R}^4$ and the two-dimensional complex Hermitian space $\mathbb{C}^2$ by identifying $(x_1, x_2, y_1, y_2) \in \mathbb{R}^4$ with the pair of complex numbers $(z_1, z_2) =$

(x_1+iy_1, x_2+iy_2). The Hermitian product is (as usual) $\langle z_1, z_2\rangle = \bar{z}_1 z_2$. Suppose that the two subspaces A^j $(j=1, 2)$ of our considerations are (one-dimensional) *complex* subspaces. The reader can easily check that this complex subspace situation precisely covers the cases where the minimum and maximum angles are equal:

$$\cos(\alpha_0) = \cos(\beta_0)$$

the common value of these cosines being just the absolute value of any (Hermitian) scalar product of two unit vectors from L^1 and L^2. Note that the modified billiard system of three disks on $\mathbb{T}^2$—mentioned in the introduction before the theorem—fits this case: the minimum and maximum angles are both $\pi/3$.

Since the rotations in L^j are unitary transformations, the generated group G must be a subgroup of the four-dimensional Lie group $U(2)$. On the other hand, since the stabilizer subgroups (in G) of the points $V \in S^3$ are at least one-dimensional, $\dim(G) \geqslant 4$, therefore G is an open subgroup of the connected group $U(2)$, showing $G = U(2)$.

The second part of this section contains the proof of Lemma 4.2. It may be a bit surprising that this proof is much simpler than that of Lemma 4.1. But one should keep in mind that in the case of Lemma 4.2 we have three islands at hand, and this fact gives us more skills for the proof.

Proof of Lemma 4.2. We can and do assume that $x^0 \in (M_3^0 \cap M_{ns}) \backslash \partial M$. For simplicity we suppose that, for some $t>0$ with $S^t x^0 \notin \partial M$, $S^{\mathbb{R}_-} x^0$ has just one island of type C^2, $S^{(0,t)} x^0$ has just one island of type C^1, and $S^{t+\mathbb{R}_+} x^0$ has just one island of type C^2. We can again apply the method of ref. 5 and, for brevity, here again we only provide the necessary vocabulary.

Set the notation $S^\tau x = x(\tau) = (Q(\tau), V(\tau))$ for an arbitrary real number τ and phase point $x = (Q, V) = (Q(0), V(0))$. Consider a very small open ball neighborhood U of x^0 in M with the following property: For every point $x \in U \cap M_3^0$ the symbolic collision structure (island structure) of x agrees with that of x^0. By the results of Section 3, the set $U \cap M_3^0 \cap M_{ns}$ is contained in the codimension-one smooth submanifold Γ of U:

$$U \cap M_3^0 \cap M_{ns} \subset \Gamma \tag{4.10}$$

where

$$\Gamma := \{x \in U : V(t) - V(0) \in \operatorname{span}(V(0), A^2) \cap L^1\} \tag{4.11}$$

Now we define the usual closed subset of Γ describing the avoiding of the first cylinder in negative time:

$$F_- := \{x \in \Gamma\text{: the negative trajectory of } x \\ \text{has no proper collision with } C^1\}$$

Since the set $U \cap M_3^0 \cap M_{ns}$ is a subset of F_-, in order to prove Lemma 4.2 it is enough to show that the closed set F_- *has no interior point* in Γ. Assume now the contrary:

The set F_- contains an open disk D in Γ. We will get a contradiction. Choose a suitably small positive number ε and set

$$\tilde{D} := \{x = (Q, V) \in U : \exists y = (Q', V') \in D \text{ such that} \\ V = V', Q - Q' \in A^2, \|Q - Q'\| < \varepsilon\} \tag{4.12}$$

It is obvious that—although the negative trajectories of the points of $\tilde{D}$ do not necessarily avoid the original cylinder C^1—they do avoid the modified first cylinder that can be obtained from C^1 by shrinking it by ε. Along the lines of the proof of Sublemma 2 of Section 5 in ref. 3, by an application of the measure-theoretic "ball avoiding lemma" for the subbilliard with the lone cylinder C^2 we get that the set $\tilde{D}$ must have measure zero. We will arrive at a contradiction (thus proving Lemma 4.2), if we prove the following:

Lemma 4.6. For every point $x = (Q, V) \in \Gamma$ the two-dimensional smooth manifold

$$\{y = (Q', V') \in U : V = V', Q - Q' \in A^2\}$$

is transversal to Γ at x.

Proof. In the spirit of Section 3 (see also the definition of Γ and the argument at the very end of Section 3) it is enough to find a vector from the set difference

$$A^2 \backslash \operatorname{span}(V, A^1)$$

But just this task was accomplished at the end of the mentioned section. Hence Lemma 4.2. ■

Singular Orbits. To end this section we turn to some thoughts about the proof of Main Lemma 2.4. Our subset C_{es} of eventually simple points is analogous to the subset C_{ed} of eventually decomposing points used in ref. 6 (and in ref. 12). For any $t_0 \in \mathbb{R}$ and any $j = 1, 2$, we can also introduce the subset

$$C_{es}(t_0, j) := \{x \in C_{es} : Q(S^{t_0 + \mathbb{R}_+} x) \cap \partial C^{3-j} = \emptyset\}$$

(No collision with C^{3-j} after t_0.) Without loss of generality we can consider the case $j = 2$. Now the definitions of y_0, $U(y_0)$, F_+, F'_+ are evidently the same as in ref. 6:

$$F_+ := \{z \in U(y_0) : Q(S^{\mathbb{R}_+} z) \cap \partial C^1 = \emptyset\}$$
$$F'_+ := \{z \in U(y_0) : Q(S^{\mathbb{R}_+} z) \cap \partial C^{1*} = \emptyset\}$$

(the star denotes the modified cylinder shrunk by some ε), whereas for $y := (Q, V)$

$$\gamma_0^s(y) := \{(Q', V) : Q' = Q + \lambda V + A^2, \lambda \in \mathbb{R}\}$$

and thus $\dim \gamma_0^s = 3$. Further, the definition of $\gamma_{\exp}^s(y)$ from Section 4 is adopted instead of the manifolds $\gamma_{1,2}^s(\cdot)$ and $\gamma_{3,4}^s(\cdot)$. Of course, $\dim \gamma_{\exp}^s = 1$. The dimension of

$$\gamma_0^{sp}(y) := \{z \in \gamma_0^s(y) : Q(z) - Q(y) \perp V(y)\}$$

is, of course, 2. The dimension of the generate

$$\gamma_g^s(y) := \bigcup_{z \in \gamma_0^{sp}(y)} \gamma_e^s(z) = \bigcup_{z \in \gamma_e^s(y)} \gamma_0^{sp}(z)$$

is then 3, as required. It is easy to see that, with these modifications in the definitions, the proof of Main Theorem 6.1 of ref. 5 can literally be copied. Hence Main Lemma 2.4. ■

ACKNOWLEDGMENT

This research was supported by the Hungarian National Foundation for Scientific Research, grant no. 1902.

REFERENCES

1. L. Bunimovich, C. Liverani, A. Pellegrinotti, and Yu. Sukhov, Special systems of hard balls that are ergodic, *Commun. Math. Phys.* **146**:357–396 (1992).
2. L. A. Bunimovich and Ya. G. Sinai, On a fundamental theorem in the theory of dispersing billiards, *Mat. Sbornik* **90**:415–431 (1973).
3. 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).

4. A. Krámli, N. Simányi, and D. Szász, A "transversal" fundamental theorem for semi-dispersing billiards, *Commun. Math. Phys.* **129**:535–560 (1990).
5. A. Krámli, N. Simányi, and D. Szász, The K-property of three billiard balls, *Ann. Math.* **133**:37-72 (1991).
6. A. Krámli, N. Simányi, and D. Szász, The K-property of four billiard balls, *Commun. Math. Phys.* **144**:107–148 (1992).
7. Ya. G. Sinai, Dynamical systems with elastic reflections, *Usp. Mat. Nauk* **25**:141–192 (1970).
8. Ya. G. Sinai, Billiard trajectories in polyhedral angle, *Usp. Mat. Nauk* **33**:231–232 (1978).
9. N. Simányi, The K-property of N billiard balls I, *Invent. Math.* **108**:521–548 (1992); II, *Invent. Math.* **110**:151–172 (1992).
10. Ya. G. Sinai and N. I. Chernov, Ergodic properties of some systems of 2-D discs and 3-D spheres, *Usp. Mat. Nauk* **42**:153–174 (1987).
11. D. Szász, Ergodicity of classical hard balls, *Physica A* **194**:86–92 (1993).
12. D. Szász, The K-property of "orthogonal" cylindric billiards, *Commun. Math. Phys.*, to appear.