On Topological Entropy of Billiard Tables with Small Inner Scatterers

Yi-Chiuan Chen*
Institute of Mathematics, Academia Sinica, Taipei 11529, Taiwan
TEL: +886 2 27851211 ext. 307, FAX: +886 2 27827432
Email: YCChen@math.sinica.edu.tw

July 9, 2009

Abstract

We present in this paper an approach to studying the topological entropy of a class of billiard systems. In this class, any billiard table consists of strictly convex domain in the plane and strictly convex inner scatterers. Combining the concept of anti-integrable limit with the theory of Lyusternik-Shnirel'man, we show that a billiard system in this class generically admits a set of non-degenerate anti-integrable orbits which corresponds bijectively to a topological Markov chain of arbitrarily large topological entropy. The anti-integrable limit is the singular limit when scatterers shrink to points. In order to get around the singular limit and so as to apply the implicit function theorem, on auxiliary circles encircling these scatterers we define a length functional whose critical points are well-defined at the anti-integrable limit and give rise to billiard orbits when the scatterers are not points. Consequently, we prove the topological entropy of the first return map to the scatterers can be made arbitrarily large provided the inner scatterers are sufficiently small.

Mathematics Subject Classification: 37B10, 37B40, 37C35, 37D50

Keywords: Topological entropy, Anti-integrable limit, Billiards

*This work was partially supported by grants NSC 95-2115-M-001-023, NSC 96-2115-M-001-007, and NSC 97-2119-M-001-008

1 Introduction and main results

Tracing back to 1970, Sinai [33] observed that parallel wavefront of rays diverges when it reflects from concave boundary, thus billiards with concave boundary potentially admit hyperbolic structure. His seminal paper established a connection between Boltzmann's ergodic hypothesis of statistical mechanics and the hyperbolicity and ergodicity of semi-dispersing billiards. For the now so-called Sinai billiard system, he proved that it has positive measure-theoretic (Kolmogorov-Sinai) entropy and is hyperbolic almost everywhere. See also [11, 18, 32, 34] and also [14, 26, 29, 30, 31] for relevant and recent results and references therein. In contrast, rays converge after reflecting from convex boundaries. If the boundary of a billiard system is strictly convex and sufficiently smooth (C^6 is sufficient), Lazutkin's result [24] on caustics showed that the system cannot be ergodic and that not almost all orbits can have non-zero Lyapunov exponents. Note that Bunimovich [6] constructed convex C^1-tables (the Bunimovich stadium billiards) with non-zero Lyapunov exponents almost everywhere. Traditionally, semi-dispersing billiards are investigated from the viewpoint of ergodic theory, while billiards in smooth convex domains are studied by means of the twist maps (see, e.g. [3, 22]).

Much less is known in the case that the boundaries of billiard tables are mixed with both concave and convex curves, for example, strictly convex billiard tables with circular inner scatterers. Foltin [15, 16] recently proved a nice result that billiard flows on strictly convex C^2-tables with sufficiently small inner disjoint circular scatterers generically possess positive topological entropy. More precisely, his result may be described as follows. Let $M \subset \mathbb{R}^2$ be the domain of billiard table, ∂M the boundary of the table, $O_1, O_2, \ldots, O_K$ the centres of the circular scatterers $B_1, B_2, \ldots, B_K$ located in the interior of M. Let ∂M be parametrised as

$$\begin{aligned}\partial M := \{\phi(\theta) = (p(\theta)\cos\theta,\ p(\theta)\sin\theta) :\\ 2\dot{p}^2 - p(\ddot{p} - p) > 0,\ 0 \le \theta < 2\pi\}.\end{aligned} \tag{1}$$

Define

$$\begin{aligned}\{(p, O) \in C^2(\mathbb{R}/(2\pi\mathbb{Z}), R_+) \times \mathbb{R}^{2K} :\ p \text{ satisfies (1)},\\ O = (O_1, \ldots, O_K),\ O_e \in \text{interior}(M)\ \forall\, e = 1, \ldots, K\}\end{aligned} \tag{2}$$

to be the space of convex billiards having inner circular scatterers endowed with an inherited product metric arising from the C^2-metric and the usual metric on $\mathbb{R}^2$.

Theorem 1.1 (Foltin [15, 16]). *There is an open and dense subset of the billiard space (2) with $K \ge 1$ in which every billiard flow has positive topological entropy provided the inner scatterers are small enough.*

Remark 1.2. Even if there is no any inner scatterer, Cheng [10] later showed that the topological entropy of a strictly convex C^3-table is generically positive. (But, for a C^1-map $\mathcal{T} : \mathcal{Z} \to \mathcal{Z}$ of a compact Riemannian manifold, its topological entropy is at most

$\dim(\mathcal{Z}) \ln \sup_{z\in\mathcal{Z}} \|D\mathcal{T}(z)\|$ [22, 36].) When the number K of the convex scatterers is greater than or equal to three, and when the so-called "no eclipse" condition is fulfilled, the topological entropy of the billiard flow is between $\ln(K-1)/\mathrm{diam}\,(B_1 \cup \ldots \cup B_K)$ and $\ln(K-1)/\min_{i\neq j} \mathrm{dist}\,(B_i, B_j)$ even if ∂M is removed [29, 35].

The essential ideas of [15, 16] are to show, with sufficiently small inner discs, for a generic billiard system in the space (2) there exist (at least) two period-2 orbits which perpendicularly collide with the billiard boundary and with the inner scatterer, then to show these two orbits admit the shift automorphism on two symbols. In a situation when these two orbits lie on a line with the centre of one of the inner discs (e.g. see Figure 1 of [16] or Figure 2(a) of this paper), Foltin also showed there exists an additional period-6 orbit, and these three orbits (two period-2, one period-6) admit a subshift of finite type of positive topological entropy on four symbols.

One of the main aims of this article is to show that Foltin's result on the positivity of the topological entropy can alternatively be understood as a property that is inherited from the anti-integrable limit [1, 2, 8, 9, 27]. Observe that the billiard systems considered by Foltin have a singular limit when the scatterers shrink to points O. This kind of singular limit or called *small-scatterer limit* has drawn increasing attention to the study of billiards, see e.g. [5, 8, 12, 13, 17, 19, 20, 28], also a study of small scatterer problem about rotation sets [4]. In the spirit of [8], we call such a limit the *anti-integrable limit*. In the limiting situation, we are interested in those orbits which start from and return back to the set O, after several bounces on the boundary ∂M. Also we want to know what happens to these orbits when a system is near the limit.

To elucidate what we mean, consider the instance in Figure 1(a). In the figure, M

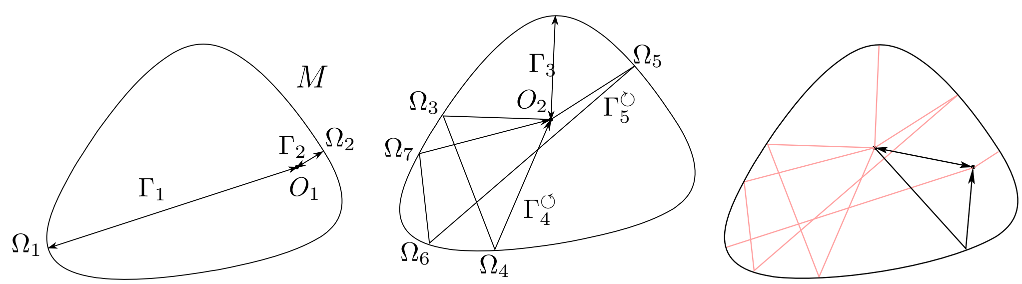

Figure 1: (a) Two 1-link basic AI-orbits Γ_1 and Γ_2 for O_1. (b) Three basic AI-orbits Γ_3, $\Gamma_4^{\circlearrowleft}$ and $\Gamma_5^{\circlearrowleft}$ for O_2 of respectively 1-link, 2-link and 3-link. (c) One 0-link and one 1-link basic AI-orbits connecting O_1 and O_2.

is a bounded domain whose boundary ∂M is a simple closed C^3-curve, and $\Gamma_1 := \overline{O_1\Omega_1} \cdot \overline{\Omega_1 O_1}$, $\Gamma_2 := \overline{O_1\Omega_2} \cdot \overline{\Omega_2 O_1}$ are product paths (or products of line segments). Suppose $\overline{O_1\Omega_1}$, $\overline{O_1\Omega_2}$ intersect perpendicularly with ∂M, and suppose Ω_1, O_1, Ω_2 are not collinear. If now the point O_1 is replaced by a small circular scatterer centred at O_1, then it is apparent that there will exist two period-2 orbits, one along Γ_1, the other along Γ_2. Moreover, using the approach of anti-integrable limit, we can show that, in general, for any sequence $\{b_i\}_{i\in\mathbb{Z}}$ with $b_i \in \{1, 2\}$ there is a unique orbit "shadowing" the product path $\cdots \Gamma_{b_{-1}} \cdot \Gamma_{b_0} \cdot \Gamma_{b_1} \cdots$ provided the circular scatterer is sufficiently small. As a consequence, the positivity of the topological entropy of the system results from the shift automorphism on two symbols. The product path $\cdots \Gamma_{b_{-1}} \cdot \Gamma_{b_0} \cdot \Gamma_{b_1} \cdots$ is called an *anti-integrable orbit* (abbreviated *AI-orbit*) of the system. Hence, the entropy of the system in this case is at least as that of the anti-integrable orbits.

We call piecewise straight paths like Γ_1 and Γ_2 1-link basic AI-orbits, which are billiard orbit segments starting from and ending at a given point-scatterer (namely the O_1 in the case of Γ_1 and Γ_2) with one bounce with ∂M. More generally, we can define n-link basic AI-orbits.

Definition 1.3. *For an integer $n \geq 1$ and for $e_1, e_2 \in \{1, \ldots, K\}$, a piecewise straight path Γ is called an n-**link basic AI-orbit** connecting two (not necessarily different) point-scatterers O_{e_1} and O_{e_2} if it is a segment of billiard orbit starting from one of these two scatterers and having exactly n number of consecutive bounces with the boundary ∂M before reaching the other one. In the case that O_{e_1}, O_{e_2} are different and Γ has no bounce (i.e. a straight line segment), we call Γ a 0-link basic AI-orbit.*

Remark 1.4. In this paper we regard point-scatterers as obstacles, thus no billiard orbits can go straight through them.

Figure 1(b) illustrates examples of 2-link and 3-link basic AI-orbits. $\Gamma_4^{\circlearrowleft}$ is a 2-link basic AI-orbit from point O_2, bouncing off ∂M at Ω_3, then at Ω_4, then back to O_2. Having such a basic AI-orbit, we can construct another basic AI-orbit $\Gamma_4^{\circlearrowright}$ which leaves O_2 for Ω_4, then bounces off Ω_3 before returning to O_2. We call these two basic AI-orbits $\Gamma_4^{\circlearrowleft}$ and $\Gamma_4^{\circlearrowright}$ *geometrically indistinct*. Two basic AI-orbits which are not geometrically indistinct are called *geometrically distinct*. It is possible that not all n-link basic AI-orbits exist for some special shapes of the boundary ∂M. When ∂M is a circle and O_1 is the only inner point scatterer and is located in the centre of the circle as depicted in Figure 2(a), there is a continuous family of 1-link basic AI-orbits, but no other n-link basic AI-orbits with $n \geq 2$. Denote the Euclidean distance between two points x and y by $h(x, y)$:

$$h(x, y) = |y - x|, \quad x, y \in \mathbb{R}^2.$$

The existence of 1-link basic AI-orbits is obvious in general since the function $h(O_e, \cdot) : \partial M \to \mathbb{R}$, $1 \leq e \leq K$, attains its global maximum and minimum. For multi-link case, we invoke the Lyusternik-Shnirel'man theory (see Proposition 3.4 and Corollary 3.6) to show the existence of n-link basic AI-orbits for all $n \geq 2$.

In [8], the author obtained a lower bound estimate of the topological entropy of a generalized Sinai billiard system. It is a Hamiltonian system on the two dimensional torus with a steep Coulomb-type repulsive potential of the form $\overline{V}_\rho(x, \epsilon) = \epsilon/(|x| - \rho/2)$ with ρ, $\epsilon > 0$, i.e. a soft scatterer, cf. [14, 30, 31]. The author showed that the lower bound can be made arbitrarily large provided that ρ and ϵ are sufficiently small. The reasons for this are because there exists a unique basic AI-orbit in the

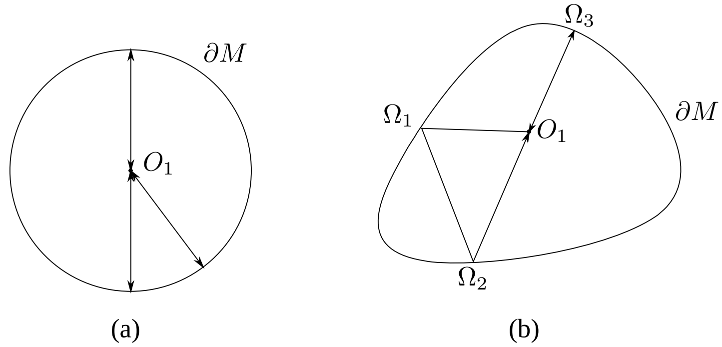

Figure 2: (a) ∂M is a circle with centre O_1. (b) $\overline{O_1\Omega_3}$ intersects perpendicularly with ∂M. Suppose Ω_3, O_1, Ω_2 are collinear, then the product paths $\overline{O_1\Omega_1} \cdot \overline{\Omega_1\Omega_2} \cdot \overline{\Omega_2 O_1}$ and $\overline{O_1\Omega_3} \cdot \overline{\Omega_3 O_1}$ are respectively 2-link and 1-link basic AI-orbits, but $\overline{O_1\Omega_1} \cdot \overline{\Omega_1\Omega_2} \cdot \overline{\Omega_2\Omega_3} \cdot \overline{\Omega_3 O_1}$ is *not* a 3-link basic AI-orbit.

limits $(\rho, \epsilon) = (0, 0)$ in any homotopy class of loops based on the center of the torus and because for sufficiently small (ρ, ϵ) there corresponds a unique Hamiltonian orbit shadows a prescribed bi-infinite chain of basic AI-orbits. Since the fundamental group of the torus is $\mathbb{Z}^2$, the lower bound of the topological entropy of the first return map of the Hamiltonian flow to a fixed cross-section in the phase space of the generalized Sinai billiard can therefore be made as large as we wish. The hard scatterer case can be modeled by considering the limit $\lim_{\epsilon \to 0} \overline{V}_\rho(\cdot, \epsilon)$, cf. [30, 31]. This limiting case describes the Sinai billiard on the two torus with a circular scatterer of diameter ρ, and in this case the lower bound goes to infinity at the rate $-2 \ln \rho + O(1)$ as ρ goes to zero. (See Remark 1.5.)) By the persistence of hyperbolic orbits under C^1-perturbations, we can conclude that the lower bound is of order $-2 \ln \rho$ for sufficiently small ρ and ϵ.

As pointed out in [33] that the Sinai billiard is equivalent to the billiard system of square table with a circular scatterer placed in the center of the square. This naturally motives us to achieve Foltin's result by the ideas used in [8]. From the fact that the system currently considered here (if having only one inner scatterer) differs from the Sinai billiard (if having a square as its table) only in the shape of the boundary curves, we actually are able to employ the method developed in [8] to the current system and obtain a much stronger result than Foltin's. (See one of the main results of this paper, Theorem 1.6.)

Remark 1.5. As a matter of fact, Chernov [11] proved that the first return map to the (hard) scatterer of the Sinai billiard has infinite topological entropy for any $0 < \rho < 1$, hence no asymptotic formulae can be correct. Notice that the entropy mentioned here is for the first return map induced by the billiard flow, not the flow itself. For the flow, if one considers the entropy as the one for the time-one map induced by the flow, then the topological entropy converges to a constant ≈ 1.526 as $\rho \to 0$ [7]. Note that the following case is also discussed in [12]: Replace the circular scatterer by a convex one of arbitrary shape, and homothetically compresses it with a scale factor of δ, then consider the limit $\delta \to 0$.

Let A_1, $A_2, \ldots, A_K$ be strictly convex domains of $\mathbb{R}^2$ depending C^3 on their diameters ρ_1, $\rho_2, \ldots, \rho_K$ and being contained inside the circular domains B_1, $B_2, \ldots, B_K$, respectively. Let their boundaries be parametrised by

$$\partial A_e := \{O_e + (g_e(\theta)\cos\theta, g_e(\theta)\sin\theta) : \\ 2\dot{g}_e^2 - g_e(\ddot{g}_e - g_e) > 0,\ 0 \le \theta < 2\pi\},$$

where $1 \le e \le K$ and g_e are positive real valued C^3-functions. (Note that O_e is the centre of the closed disc B_e for each e, and that g_e is required to depend C^3 on ρ_e.) Similar to the circular-scatterer case, we define by $(\mathcal{M}, \mathcal{O})$ the space of strictly convex C^3 billiard tables with strictly convex inner scatterers $A_1, \ldots, A_K$:

$$(\mathcal{M}, \mathcal{O}) := \{(p, O) \in C^3(\mathbb{R}/(2\pi\mathbb{Z}), \mathbb{R}_+) \times \mathbb{R}^{2K} :\ p \text{ satisfies (1)}, \\ O = (O_1, \ldots, O_K),\ O_e \in \text{interior}(M)\ \forall\, e = 1, \ldots, K\}$$

endowed with an inherited product metric arising from the C^3-metric and the usual metric on $\mathbb{R}^2$. Because ∂M is determined by p, instead of (p, O) we shall use (M, O) to represent an element of $(\mathcal{M}, \mathcal{O})$.

The billiard system induces a *billiard collision map* on the compact manifold $\partial(M \setminus \{A_1, \ldots, A_K\}) \times [-\pi/2, \pi/2]$,

$$(\omega_i, \lambda_i) \mapsto (\omega_{i+1}, \lambda_{i+1}), \qquad i \in \mathbb{Z}, \tag{3}$$

where $\ldots, \omega_{-1}, \omega_0, \omega_1, \ldots$ is a sequence of consecutive collision points on the boundary $\partial(M \setminus \{A_1, \ldots, A_K\})$, and $\lambda_i \in [-\pi/2, \pi/2]$ is the incidence angle when the particle collides with the boundary at ω_i, measured from the particle's velocity to the outward normal of the boundary. Note that, except on a subset of measure zero where singularity occurs (corresponding to the tangential collision of the billiard particle with the scatterers), the billiard map just defined by (3) is continuous, and that if $(\omega, \lambda) \in \partial M \times \{\pm\pi/2\}$ then the billiard orbit at ω is tangent to the outer boundary ∂M and (ω, λ) is a fixed point of the billiard map.

By neglecting collisions occurred on the boundary ∂M of the billiard table, the billiard system also induces a map, called the *first return map to the scatterers*, on the compact manifold $\partial(\bigcup_{e=1}^{K} A_e) \times [-\pi/2, \pi/2]$,

$$(\psi_i, \alpha_i) \mapsto (\psi_{i+1}, \alpha_{i+1}), \qquad i \in \mathbb{Z}, \tag{4}$$

where $\ldots, \psi_{-1}, \psi_0, \psi_1, \ldots$ are consecutive collision points on the boundary $\partial(\bigcup_{e=1}^{k} A_e)$ of the scatterers, and every $\alpha_i \in [-\pi/2, \pi/2]$ is the incidence angle when the billiard particle first returns to and collides with the boundary at ψ_i, measured from the particle's velocity to the outward normal of the boundary. One major difference between the collision map (3) and the first return map (4) is that the first return map may not be defined everywhere, since there may be points that never return. This also results in a fact that the "return time" of some billiard orbits may be arbitrarily long.

Theorems 1.6 and 1.11 below are the main theorems of this paper.

Theorem 1.6. *For any positive real number χ, there exists an open and dense subset of $(\mathcal{M}, \mathcal{O})$ with $K \geq 1$ in which the first return map defined by (4) has topological entropy at least χ provided the strictly convex inner scatterers are sufficiently small.*

The tool that we measure the entropy in the above theorem basically relies on Theorem 1.11.

Remark 1.7. In Theorem 1.6, how small the scatterers should be vary from system to system. We do not have a uniform lower bound ρ_0 so that if all of the diameters $\rho_1, \ldots, \rho_K$ of the inner scatterers are smaller than ρ_0, then in the open and dense subset every map defined by (4) has topological entropy at least χ. Think about the following situation (cf. the paragraph below Theorem 4.1 in [16]): For a billiard system having a circular domain of diameter one as its table and a disc of diameter ρ as its only inner scatterer, can we fix a small enough ρ so that the topological entropy of the first return map is not less than $\ln 100$ no matter how close the scatterer to the boundary of the table is?

Remark 1.8. Our proof of Theorem 1.6, which is located in Section 6, will not imply the topological entropy of the billiard collision map (3) can be made arbitrarily large even though the inner scatterers are small. (We prove this remark in Section 6.)

Given a billiard system (M, O), we let $\{U_1, U_2, \ldots, U_K\}$ be the set of closed discs of fixed diameter R centred at O_1, $O_2, \ldots, O_K$ respectively such that $A_e \subseteq B_e \subset U_e$ for all $e = 1, \ldots, K$. We assume R is sufficiently small so that these discs U_e do not overlap and are contained in the interior of M. Then the billiard orbits generated by bouncing off the boundary of the inner scatterers A_e and the billiard table M may induce a sequence of pairs on $(\bigcup_{e=1}^{K} \partial U_e)^2$ in the following way. By neglecting collisions occurred on the boundary ∂M, suppose a billiard orbit possesses successive collision points $\ldots, \psi_{-1}, \psi_0, \psi_1, \ldots$ with $\psi_i \in \partial A_{e_i}$ and $e_i \in \{1, \ldots, K\}$ for every integer i. Then, travelling from ψ_{i-1} to ψ_i, the orbit must leave the circular domain $U_{e_{i-1}}$ from a unique point, say y_{i-1}, on $\partial U_{e_{i-1}}$, and must arrive at a unique point, say x_i, on the boundary ∂U_{e_i} of U_{e_i} before reaching ψ_i. In this way, we obtain a sequence of pairs on $(\bigcup_{e=1}^{K} \partial U_e)^2$ as

$$\{\ldots, (y_{i-1}, x_i), (y_i, x_{i+1}), \ldots\}, \qquad i \in \mathbb{Z}. \tag{5}$$

Remark 1.9. We explain the reason and purpose for considering and constructing the sequence of pairs on the auxiliary circles ∂U_e (cf. the *transparent walls* used in [12, 25]). These are because the billiard system is no longer a dynamical system when one or more of the inner scatterers shrink to points. Think about the question: What direction will a billiard particle reflect to when it hits a scatterer of zero diameter? For simplicity, let us assume that there is only one scatterer and the scatterer is a disc B of diameter ρ. Then, for $\rho \neq 0$, (3) is a symplectic map on $(\partial M \cup \partial B) \times [-\pi/2, \pi/2]$. At the limit $\rho \to 0$ ("the anti-integrable limit"), however, a severe problem occurs: ∂B collapses to a point. Due to this, we instead concentrate on the behaviour about how the billiard orbits intersect with the fixed concentric circle ∂U by considering sequences of pairs defined by (5) on $\partial U \times \partial U$. As anticipated, when ρ is zero, the billiard system loses its dynamics and the sequences of pairs cannot be defined. However, since the domain $\partial U \times \partial U$ is fixed, does not change as ρ does, we shall see in Theorem 2.4 that there exists a sub-domain in $\partial U \times \partial U$ on which the sequences of pairs have well defined limiting behaviour as $\rho \to 0$.

What is a billiard orbit when $\rho = 0$?

Definition 1.10. *Given a sequence of n_i-link basic AI-orbits $\{\Gamma_{b_i}\}$ such that Γ_{b_i} starts from $O_{e_{i-1}}$ and ends at O_{e_i} for every i, define*

$$\begin{aligned} y^\dagger_{i-1} &:= \Gamma_{b_i} \pitchfork \partial U_{e_{i-1}}, \\ x^\dagger_i &:= \Gamma_{b_i} \pitchfork \partial U_{e_i}, \end{aligned}$$

*where the symbol $\pitchfork$ means "perpendicular" intersection. See Figure 3(b). If $x^\dagger_i$, O_{e_i}, $y^\dagger_i$ are not collinear for all i and if $\sup_{i\in\mathbb{Z}} n_i = N$, we call the bi-infinite product $\cdots \cdot \Gamma_{b_{i-1}} \cdot \Gamma_{b_i} \cdot \Gamma_{b_{i+1}} \cdot \cdots$ an N-**AI-orbit**. (An n-link basic AI-orbit or N-AI-orbit is said to be* **non-degenerate** *if some non-degeneracy condition is satisfied, see Definition 3.2.)*

Another way to define an N-AI-orbit is that it is an infinite path which joins point-scatterers to point-scatterers with at most N number of bounces off ∂M between two point-scatterers and forbids going straight through any point-scatterer (the path must change direction when it meets with a point-scatterer).

Let $C \geq 1$ be an integer and Σ_{2C} be the space of bi-infinite sequences $\{w_i\}$ consisting of $2C$ number of symbols $w_i \in \{1, 2, \ldots, C, -1, -2, \ldots, -C\}$ and let $\widetilde{\Sigma}_{2C}$ be the subspace

$$\widetilde{\Sigma}_{2C} := \{\{w_i\} \in \Sigma_{2C} : \ w_{i+1} \neq -w_i \ \forall \ i \in \mathbb{Z}\}. \tag{6}$$

Theorem 1.11. *Let $N \geq 1$.*
(i) There exists an open and dense subset of $(\mathcal{M}, \mathcal{O})$ in which every billiard system possesses a set of non-degenerate N-AI-orbits which corresponds bijectively to $\widetilde{\Sigma}_{2C}$ with C an integer satisfying $[\frac{3N}{2}] \leq C \leq 2N - 1$. ($[\frac{3N}{2}]$ stands for the integer part of $\frac{3N}{2}$.)
(ii) If $\tilde{z}^\dagger = \{\ldots, (\tilde{y}^\dagger_{i-1}, \tilde{x}^\dagger_i), \ldots\}$ and $z^\dagger = \{\ldots, (y^\dagger_{i-1}, x^\dagger_i), \ldots\} \neq \tilde{z}^\dagger$ are determined via Definition 1.10 by any two of these non-degenerate N-AI-orbits , then

$$\sup_{i\in\mathbb{Z}} \left\{ |\tilde{x}^\dagger_i - x^\dagger_i|_{\partial U}, \ |\tilde{y}^\dagger_i - y^\dagger_i|_{\partial U} \right\} > c$$

for some positive constant c, independent of these N-AI-orbits, where $|\cdot - \cdot|_{\partial U}$ measures the least arc length between two points on ∂U_e.
(iii) Let $m \geq 0$. The maximum cardinality of subsets of these non-degenerate N-AI-orbits from which

$$\max_{0\leq i<m} \left\{ |\tilde{y}^\dagger_{i-1} - y^\dagger_{i-1}|_{\partial U}, \ |\tilde{x}^\dagger_i - x^\dagger_i|_{\partial U} \right\} > c \tag{7}$$

for any $\tilde{z}^\dagger$ and $z^\dagger \neq \tilde{z}^\dagger$ and the same c in (ii) is $2C(2C-1)^{m-1}$.

Recall that the topological entropy $h_{\rm top}$ of a continuous map $\mathcal{T}$ of a compact metric space $\mathcal{Z}$ with metric d is given by the formula (see, e.g. [22, 36] for more details)

$$h_{\rm top} = \lim_{\epsilon\to 0}\limsup_{m\to\infty}\frac{1}{m}\ln\#(m,\epsilon), \tag{8}$$

where $\#(m,\epsilon)$ is the maximum of cardinalities of (m,ϵ)-separated sets for $\mathcal{T}$. That is, one is able to find at most $\#(m,\epsilon)$ number of points $z^1, z^2, \ldots, z^{\#(m,\epsilon)}$ in $\mathcal{Z}$ such that $\max_{0\le n<m} d(\mathcal{T}^n(z^i), \mathcal{T}^n(z^j)) > \epsilon$ for any $i \neq j$. In the case that a subshift of finite type (or called topological Markov chain) can be embedded in $\mathcal{Z}$, one can conclude the topological entropy of the map $\mathcal{T}$ is at least as large as that of the subshift. If Σ_C is the space of bi-infinite sequences consisting of C number of symbols, then $\sigma|\Sigma_C$ has topological entropy $\ln C$, where σ is the shift automorphism; if $\hat{\Sigma}_C \subset \Sigma_C$ is a subshift of finite type, then the topological entropy of $\sigma|\hat{\Sigma}_C$ is equal to

$$\limsup_{m\to\infty}\frac{1}{m}\ln\#(m), \tag{9}$$

where $\#(m)$ denotes the number of words of length m in $\hat{\Sigma}_C$. Easy calculation shows that the number of words of length m in $\widetilde{\Sigma}_{2C}$ defined in (6) is

$$\#(m) = 2C(2C-1)^{m-1},$$

thus, by means of formula (9), the topological entropy of $\sigma|\widetilde{\Sigma}_{2C}$ is $\ln(2C-1)$. In our proof of Theorem 1.6, the above number $\#(m)$ will be used to achieve a lower bound estimate of the maximal cardinality of (m,ϵ)-separated sets for the first return map (4). Although the first return map is not everywhere continuous, it is valid to use formula (8) in our discussion. This is because our concern is with a lower bound of the entropy and we shall only apply the formula to neighbourhoods of certain orbits on which the first return map is continuous.

This article is organised in the following way. In the next section, we define a function $F(\cdot,\rho)$ in a Banach space which is jointly C^1 in its variable and parameter $\rho = (\rho_1, \ldots, \rho_K)$ and the zeros of which will give rise to billiard orbits. In particular, all zeros correspond to AI-orbits when $\rho = 0$. Our exposition is to construct the function $F(\cdot,\rho)$ by considering the case that the scatterers are discs rather than directly considering the case of convex scatterers. The ideas, methods, and results in both cases are the same, but the former is descriptively and intuitively simpler. In Section 3, we prove that the zeros of $F(\cdot,0)$ are generically non-degenerate, therefore, in Section 6 we can apply the implicit function theorem to find zeros of $F(\cdot,\rho)$ for small ρ. In section 4, we show that the construction of the function $F(\cdot,\rho)$ in Section 2 is valid also in the convex scatterer case. Section 5 is devoted to the details to be used to analyse the function $F(\cdot,\rho)$.

2 Shadowing broken billiard orbits

Firstly let us assume that all the convex scatterers A_e, $e = 1, \ldots, K$, are circular and of diameters ρ_e, and use B_e to represent them, namely, $A_e \equiv B_e$. (Throughout this paper we always explicitly use B_e instead of A_e to represent and so as to emphasize circular scatterers.) Let U_e be concentric circular domains containing circular scatterers B_e, and let $\{U_{e_i}\}_{i\in\mathbb{Z}}$ be such a sequence that $U_{e_i} \in \{U_1, \ldots, U_K\}$ for every $i \in \mathbb{Z}$. We assume that each U_e has diameter R with

$$R > \max_{1\leq e\leq K}\{\rho_e\}.$$

Given sufficiently close two points x_i and y_i belonging to ∂U_{e_i}, it is obvious that there is a segment of a unique billiard orbit entering ∂U_{e_i} at x_i, bouncing off the scatterer B_{e_i} at Ψ_i, then leaving ∂U_{e_i} at y_i (see Figure 3(a)). The length of this orbit segment is $h(x_i, \Psi_i) + h(\Psi_i, y_i)$. Given another point x_{i+1} belonging to $\partial U_{e_{i+1}}$, assume there exists a segment of an orbit connecting up y_i with x_{i+1} such that this segment has at most N number of bounces with ∂M and does not hit $B_1, \ldots, B_K$, see Figure 3(a). Let the length of this segment be defined by $h^*(y_i, x_{i+1})$. Because the location of Ψ_i depends on (x_i, y_i, ρ_{e_i}), we can define another two functions h^- and h^+ by

$$h^-(x_i, y_i, \rho) \quad := \quad h(\Psi_i, y_i), \tag{10}$$

$$h^+(x_i, y_i, \rho) \quad := \quad h(x_i, \Psi_i). \tag{11}$$

Recall that $\rho = (\rho_1, \ldots, \rho_K)$ and notice that $\rho_{e_i} \in \{\rho_1, \ldots, \rho_K\}$ $\forall\, i \in \mathbb{Z}$.

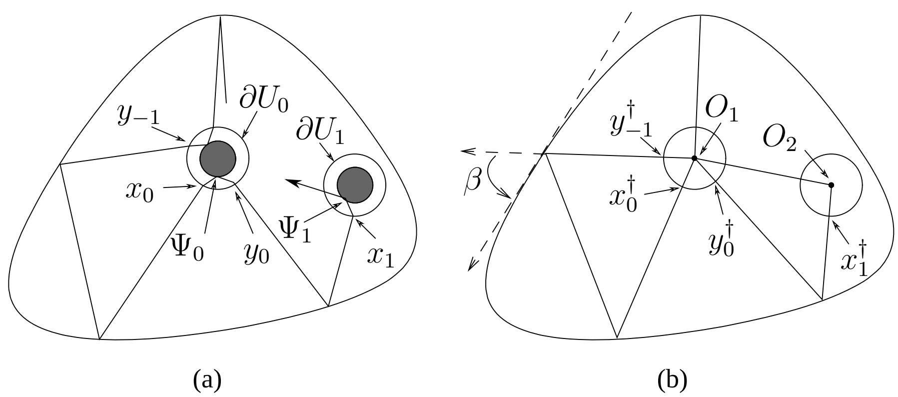

Figure 3: (a) Broken orbits determined by x_i and y_i. (b) One 0-link basic AI-orbit connecting O_1 and O_2, two 1-link basic AI-orbits (one connecting O_1 to itself, the other connecting to O_2), and one 2-link basic AI-orbit connecting O_1 to itself.

Suppose we have a sequence of pairs of points x_i and y_i, $i \in \mathbb{Z}$, such that x_i is connected backwards to y_{i-1} and forwards to y_i by segments of orbits as described in the preceding paragraph. Gluing together these segments of orbits, we get a *broken billiard orbit* with broken points x_i and y_i. If there is no velocity discontinuity occurring at every broken point (i.e. the broken points are not broken), then the broken billiard orbit is a true orbit. In what follows we define a map and show that if $\{(y_{i-1}, x_i)\}_{i\in\mathbb{Z}}$ is a zero of such map then all x_i's and y_i's are not broken.

Assume functions h^-, h^*, h^+ are sufficiently smooth and well-defined (i.e. single valued) on small open subsets of ∂U_{e_i}'s, then let us define a map $F(\cdot, \rho)$ on an open subset Z of $\prod_{i\in\mathbb{Z}}(\partial U_{e_{i-1}}\times\partial U_{e_i})$ by

$$F: Z\times[0,R)^K\to l_\infty,\qquad (z,\rho)\mapsto\{F_i(z,\rho)\}_{i\in\mathbb{Z}},\tag{12}$$

where $\rho=(\rho_1,\dots,\rho_K)$,

$$\begin{aligned} z=\{z_i\}_{i\in\mathbb{Z}} &:= \{(y_{i-1},x_i)\}_{i\in\mathbb{Z}},\\ F_i(z,\rho) &:= D_{z_i}(h^+(x_{i-1},y_{i-1},\rho_{e_{i-1}})+h^-(x_{i-1},y_{i-1},\rho_{e_{i-1}})\\ &\qquad +h^*(y_{i-1},x_i)+h^+(x_i,y_i,\rho_{e_i})+h^-(x_i,y_i,\rho_{e_i})),\end{aligned}$$

and l_∞ is the subspace of $(\mathbb{R}^2\times\mathbb{R}^2)^{\mathbb{Z}}$ with the bounded sup norm. (We also endow the Cartesian product space $\prod_{i\in\mathbb{Z}}(\partial U_{e_{i-1}}\times\partial U_{e_i})$ with the bounded sup norm.) A noteworthy fact is the following.

Proposition 2.1. *Assume all $\rho_1,\dots,\rho_K$ are non-zero. Then a zero $\{(y_{i-1},x_i)\}_{i\in\mathbb{Z}}$ of $F(\cdot,\rho)$ corresponds to a unique orbit connecting points in the order* $\dots$, x_0, Ψ_0, y_0, x_1, Ψ_1, y_1, $\dots$.

Proof. The proof of the proposition (and some other results in this paper) relies on the very useful Lemma 2.2 below. Such x_i's and y_i's in the proposition mean for each i that $F_i(\{(y_{i-1},x_i)\}_i,\rho)=0$. In other words,

$$D_{y_{i-1}}(h^+(x_{i-1},y_{i-1},\rho_{e_{i-1}})+h^-(x_{i-1},y_{i-1},\rho_{e_{i-1}})+h^*(y_{i-1},x_i)) = 0,\tag{13}$$

$$D_{x_i}(h^*(y_{i-1},x_i)+h^+(x_i,y_i,\rho_{e_i})+h^-(x_i,y_i,\rho_{e_i})) = 0.\tag{14}$$

By (10), we get

$$\begin{aligned}&D_{y_{i-1}}h^-(x_{i-1},y_{i-1},\rho_{e_{i-1}})\\ &\quad=D_{y_{i-1}}h(\Psi_{i-1},y_{i-1})+D_{\Psi_{i-1}}h(\Psi_{i-1},y_{i-1})D_{y_{i-1}}\Psi_{i-1}(x_{i-1},y_{i-1},\rho_{e_{i-1}}),\\ &D_{x_i}h^-(x_i,y_i,\rho_{e_i})\\ &\quad=D_{\Psi_i}h(\Psi_i,y_i)D_{x_i}\Psi_i(x_i,y_i,\rho_{e_i});\end{aligned}$$

by (11), we have

$$\begin{aligned}&D_{y_{i-1}}h^+(x_{i-1},y_{i-1},\rho_{e_{i-1}})\\ &\quad=D_{\Psi_{i-1}}h(x_{i-1},\Psi_{i-1})D_{y_{i-1}}\Psi_{i-1}(x_{i-1},y_{i-1},\rho_{e_{i-1}}),\\ &D_{x_i}h^+(x_i,y_i,\rho_{e_i})\\ &\quad=D_{x_i}h(x_i,\Psi_i)+D_{\Psi_i}h(x_i,\Psi_i)D_{x_i}\Psi_i(x_i,y_i,\rho_{e_i}).\end{aligned}$$

Thus, $F_i(\{(y_{i-1},x_i)\}_i,\rho)=0$ if and only if

$$\begin{aligned}&D_{y_{i-1}}h(\Psi_{i-1},y_{i-1})+D_{y_{i-1}}h^*(y_{i-1},x_i)\\ &\quad= -D_{y_{i-1}}\Psi_{i-1}(x_{i-1},y_{i-1},\rho_{e_{i-1}})\left(D_{\Psi_{i-1}}(h(\Psi_{i-1},y_{i-1})+h(x_{i-1},\Psi_{i-1}))\right),\end{aligned}\tag{15}$$

$$\begin{aligned}&D_{x_i}h^*(y_{i-1},x_i)+D_{x_i}h(x_i,\Psi_i)\\ &\quad= -D_{x_i}\Psi_i(x_i,y_i,\rho_{e_i})\left(D_{\Psi_i}(h(x_i,\Psi_i)+h(\Psi_i,y_i))\right).\end{aligned}\tag{16}$$

By Lemma 2.2 and by the law of reflection at Ψ_i, the right hand sides of both the equalities vanish automatically. Thus the left hand sides also vanish. This means no velocity discontinuity occurs at x_i or y_i for every i. $\square$

Lemma 2.2. *Let x, $\phi \in \mathbb{R}^n$ and $h(x,\phi) = |\phi - x|$. Then,*

$$\begin{aligned} D_\phi h(x,\phi) &= \frac{(\phi-x)^T}{|\phi-x|}, \\ D^2_{x\phi} h(x,\phi) &= -\frac{1}{|\phi-x|^3}(\phi-x)(\phi-x)^T - \frac{1}{|\phi-x|} I, \\ D^2_{\phi\phi} h(x,\phi) &= -\frac{1}{|\phi-x|^3}(\phi-x)(\phi-x)^T + \frac{1}{|\phi-x|} I \end{aligned}$$

in which I is the n by n identity matrix, and $(\phi - x)^T$ means the transpose of the n-vector $\phi - x$.

Hence, the problem of finding billiard orbits reduces to finding zeros of $F(\cdot, \rho)$. In particular, we need to verify that $F(\cdot, \rho)$ is indeed well defined. Theorem 2.4 below shows that there exist a subset Z of $\prod_{i\in\mathbb{Z}}(\partial U_{e_{i-1}} \times \partial U_{e_i})$ and a positive constant ρ_0 so that F is continuously differentiable on $Z \times [0, \rho_0)^K$.

Remark 2.3. Another angle to look at the map $F(\cdot, \rho)$ is to consider periodic orbits. If a billiard orbit is periodic, it repeats the same orbit points after a certain number of bounces, say m bounces ($m \geq 2$), namely

$$\Psi_{i+m} = \Psi_i \qquad \forall i \in \mathbb{Z},$$

or

$$y_{i+m} = y_i \quad \text{and} \quad x_{i+m} = x_i \; \forall i.$$

In this case a zero z of $F(\cdot, \rho)$ in proposition 2.1 can be obtained by finding a critical point of a length function:

$$D_{\tilde{z}} W_m(\tilde{z}, \rho) = 0$$

where $W_m(\tilde{z}, \rho)$ is the sum

$$W_m(\tilde{z}, \rho) = \sum_{i=1}^{m} h^-(\tilde{x}_{i-1}, \tilde{y}_{i-1}, \rho) + h^*(\tilde{y}_{i-1}, \tilde{x}_i) + h^+(\tilde{x}_i, \tilde{y}_i, \rho)$$

and $\tilde{z} \in Z_m$ with Z_m being the finite dimensional subspace of Z that $z_{i+m} = z_i$ for all $i \in \mathbb{Z}$. See more related examples in [8, 22, 23, 34].

Now consider a special case. Suppose $\cdots \Gamma_{b_{i-1}} \cdot \Gamma_{b_i} \cdot \Gamma_{b_{i+1}} \cdots$ is such an AI-orbit that $x_i^\dagger \equiv y_i^\dagger$ for all i. Then when the point-scatterers are fatten to small circular scatterers, it is apparent that there is a unique billiard orbit entering and leaving ∂U_{e_i} both at $x_i^\dagger$ for every i, with a perpendicular collision with B_{e_i}. If Γ_{b_i} is of n_i-link, then it is also easy to see that this orbit is of period-$2(n_i + 1)$ or of period-$(n_i + 1)$ and lies exactly on the closure of $\Gamma_{b_i} \setminus ((\Gamma_{b_i} \cap B_{e_{i-1}}) \cup (\Gamma_{b_i} \cap B_{e_i}))$. Hence we conclude a fact that the sequence $\{\ldots, (x_{i-1}^\dagger, x_i^\dagger),\ (x_i^\dagger, x_{i+1}^\dagger), \ldots\}$ is a solution of $F(z, \rho) = 0$ for every sufficiently small given ρ. In fact, we have

Theorem 2.4. *Suppose every n-link basic AI-orbit is non-degenerate for $n \leq N$. Given an N-AI-orbit determined by*

$$z^\dagger = \{(y^\dagger_{i-1}, x^\dagger_i)\}_{i\in\mathbb{Z}},$$

then on each ∂U_{e_i} there exist subsets Δ_{y_i} containing $y^\dagger_i$, Δ_{x_i} containing $x^\dagger_i$, and exists $\rho_0 > 0$, independent of the N-AI-orbit, such that Δ_{y_i} and Δ_{x_i} are topologically open intervals and that F is continuously differentiable on the subset $\prod_{i\in\mathbb{Z}}(\Delta_{y_{i-1}} \times \Delta_{x_i}) \times [0, \rho_0)^K$. Moreover, $F(\cdot, \rho)$ has a unique simple zero on the subset. In particular, $F(z^\dagger, 0) = 0$.

Remark 2.5. Because $F(z^\dagger, 0) = 0$ and F is C^1, we can choose a constant $C > 0$ and sufficiently small Δ_{y_i}, Δ_{x_i} and ρ_0 so that $|F_i(z, \rho)| < C$ for every $(z, \rho) \in \prod_{i\in\mathbb{Z}}(\Delta_{y_{i-1}} \times \Delta_{x_i}) \times [0, \rho_0)^K$ and $i \in \mathbb{Z}$. This means that F has uniformly bounded components.

Remark 2.6. With the notation $T(\partial U_e) = \bigcup_{x\in\partial U_e} T_x(\mathbb{R}^2)$ for the tangent bundle, the derivative $D_zF(\cdot, \rho)$ is a tangent map from the subset $\prod_{i\in\mathbb{Z}}(T(\Delta_{y_{i-1}}) \times T(\Delta_{x_i}))$ of $\prod_{i\in\mathbb{Z}}(T(\partial U_{e_{i-1}}) \times T(\partial U_{e_i}))$ into $(T(\mathbb{R}^2) \times T(\mathbb{R}^2))^{\mathbb{Z}}$. Then, $D_zF(z, \rho)$ is a linear map from $\prod_{i\in\mathbb{Z}}(T_{y_{i-1}}(\Delta_{y_{i-1}}) \times T_{x_i}(\Delta_{x_i}))$ to $(\mathbb{R}^2 \times \mathbb{R}^2)^{\mathbb{Z}}$. Since Δ_{y_i} and Δ_{x_i} are homeomorphic to open intervals, $D_zF(z, \rho)$ can be treated as a continuous family (with respect to both z and ρ) of linear maps from $(\mathbb{R}^2 \times \mathbb{R}^2)^{\mathbb{Z}}$ to $(\mathbb{R}^2 \times \mathbb{R}^2)^{\mathbb{Z}}$.

In order to prove Theorem 2.4, in the next section we define and show the existence of non-degenerate basic AI-orbits needed for the assumption of Theorem 2.4. In Sections 4, 5 and 6, we show the non-degeneracy of basic AI-orbits implies two facts: one is the existence of such Δ_{x_i} and Δ_{y_i}, and the other one is that the zero $z^\dagger$ is simple and unique. These two facts are proved in Proposition 6.1 and Lemma 6.2. To see that $z^\dagger$'s are solutions for $F(z, 0) = 0$, note by our construction that

$$D_{x_i}\Psi_i(x_i, y_i, 0) = D_{y_i}\Psi_i(x_i, y_i, 0) = 0 \qquad \forall\ x_i,\ y_i$$

and that $\Psi_i(x_i, y_i, 0) \equiv O_{e_i}$. Therefore, the right hand sides of equalities (15) and (16) are both zero when ρ is zero. The left hand sides also vanish, because $x^\dagger_i$ as well as $y^\dagger_i$ come from basic AI-orbits and no velocity discontinuity occurs over there.

In consequence, the zero of $F(\cdot, \rho)$, denoted by

$$z^*(\rho) = \{(y^*_{i-1}, x^*_i)\}_i,$$

forms a C^1-family as ρ varies, also every N-AI-orbit can be continued to an orbit which intersects ∂U_{e_i} at y^*_i and x^*_i. Because two different AI-orbits result in two different z^*'s, the positiveness of the topological entropy is a corollary of the above theorem if the considered system possesses a subset of AI-orbits which forms a Markov chain of positive topological entropy. This is the issue handled in Theorem 1.11.

So far the results of Proposition 2.1 and Theorem 2.4 are for circular scatterers, but actually they are still valid if we replace circular scatterers by strictly convex ones. In Section 4, we give a detailed investigation in this regard.

3 Generic existence of non-degenerate AI-orbits

In this section, we assume $n \geq 2$ and assume the arc-length of ∂M is normolised to one. Let

$$Q^n := \{(\phi_1, \ldots, \phi_n) \in (\partial M)^n : \ \phi_j \neq \phi_{j+1} \, \forall j = 1, \ldots, n-1\}.$$

Definition 3.1. *Let O_{e_1} and O_{e_2} be two points inside M, define the function $h_n : Q^n \to \mathbb{R}$ by*

$$h_n(\phi_1, \ldots, \phi_n) := h(O_{e_1}, \phi_1) + \sum_{j=1}^{n-1} h(\phi_j, \phi_{j+1}) + h(\phi_n, O_{e_2}). \tag{17}$$

It is easy to see that an n-link basic AI-orbit corresponds to a critical point of h_n, conversely, a critical point of h_n gives rise to an n-link basic AI-orbit provided that line segments $\overline{\phi_j \phi_{j+1}}$ do not intersect points O_e for all $j = 1, \ldots, n-1$ and $e = 1, \ldots, K$ and that $\overline{O_{e_1}\phi_1} \cap \{O_1, \ldots, O_K\} = O_{e_1}$ and $\overline{O_{e_2}\phi_n} \cap \{O_1, \ldots, O_K\} = O_{e_2}$.

Definition 3.2. *An n-link basic AI-orbit $\Gamma = \overline{O_{e_1}\phi_1} \cdot \overline{\phi_1\phi_2} \cdot \cdots \cdot \overline{\phi_n O_{e_2}}$ is said to be* **non-degenerate** *if $(\phi_1, \ldots, \phi_n)$ is a non-degenerate critical point of h_n. An AI-orbit $\cdots \Gamma_{b_i} \cdot \Gamma_{b_{i+1}} \cdots$ is called* **non-degenerate** *if Γ_{b_i} is a non-degenerate basic AI-orbit for every i.*

In order to obtain a critical point, we utilise the method of proving the Poincaré-Birkhoff theorem for billiards in [23], and define the following ϵ-conditions. (Notice that our ϵ-conditions are different from the ones in [23].)

Definition 3.3. *A point $(\phi_1, \ldots, \phi_n)$ in Q^n is said to satisfy the* ϵ**-conditions** *if for every $j \in \{1, \ldots, n-1\}$*

- $|\phi_{j+1} - \phi_j|_{\partial M} > \epsilon$*;*
- $|\phi_{j+2} - \phi_j|_{\partial M} > 3\epsilon$ *if ϕ_{j+1} lies on the least arc bounded by ϕ_j and ϕ_{j+2};*
- $|\phi_{j+3} - \phi_j|_{\partial M} > 3^2\epsilon$ *if ϕ_{j+1}, ϕ_{j+2} lie on the least arc bounded by ϕ_j and ϕ_{j+3} in the order ϕ_j, ϕ_{j+1}, ϕ_{j+2}, ϕ_{j+3} for one of the directions on the boundary ∂M;*

$$\vdots$$

- $|\phi_n - \phi_j|_{\partial M} > 3^{n-j-1}\epsilon$ *if ϕ_{j+1}, $\phi_{j+2}, \ldots, \phi_{n-1}$ lie on the least arc bounded by ϕ_j and ϕ_n in the order ϕ_j, $\phi_{j+1}, \ldots, \phi_n$ for one of the directions on ∂M.*

In the definition, $|\phi'' - \phi'|_{\partial M}$ denotes the least arc-length bounded by the two points ϕ'' and ϕ' on ∂M. Similar to what is performed in [23], we "trim" the domain Q^n by defining

$$Q^{n,\epsilon} := \{(\phi_1, \ldots, \phi_n) \in Q^n : \ (\phi_1, \ldots, \phi_n) \text{ satisfies the } \epsilon\text{-conditions}\}.$$

It is not difficult to see that h_n is C^3 on $Q^{n,\epsilon}$ and that $Q^{n,\epsilon}$ is homeomorphic to the product of the circle and the $(n-1)$-dimensional open disc. The reason for the latter is because if arbitrarily choose a ϕ_1 in ∂M then ϕ_2 belongs to ∂M minus a small open arc containing ϕ_1, and ϕ_3 belongs to ∂M minus a small open arc containing ϕ_2, etc., and ϕ_n belongs to ∂M minus a small open arc containing ϕ_{n-1}.

Proposition 3.4. *The function h_n attains at least two critical values on $Q^{n,\epsilon}$ for sufficiently small ϵ. At least one of the two critical values is a maximum, but not all of the critical values of h_n on $Q^{n,\epsilon}$ are isolated maxima.*

Proof. The proof relies on the following proposition.

Proposition 3.5 (Proposition 2.2 in [23]). *Suppose ϕ_0 lies on the least arc between $\phi^{'}$ and $\phi^{''}$ on ∂M satisfying $|\phi^{'} - \phi_0|_{\partial M} = \delta$ and $|\phi^{''} - \phi^{'}|_{\partial M} \geq 3\delta$. Let $\mathbf{t}$ be the unit tangent vector at ϕ_0 in the direction from ϕ_0 to $\phi^{''}$ along the least arc (see Figure 4). Then there exists $\delta_0 > 0$ such that*

$$\langle D_{\phi_0}(h(\phi^{'}, \phi_0) + h(\phi_0, \phi^{''})), \mathbf{t} \rangle > 0$$

provided $0 < \delta < \delta_0$.

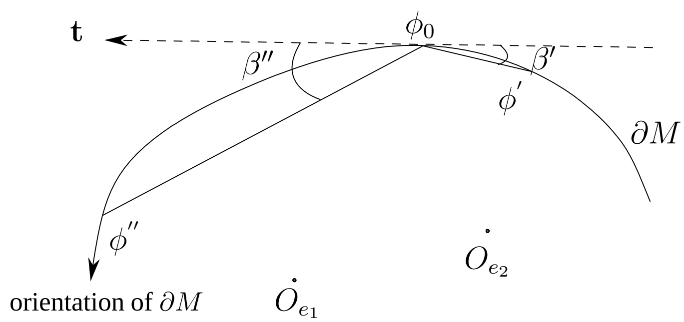

Figure 4: $|\phi^{'} - \phi_0| + |\phi_0 - \phi^{''}|$ will become larger if ϕ_0 shifts a bit in the direction to $\phi^{''}$ along the boundary arc.

The reason for Proposition 3.5 is due to a fact that the angle β' between the tangent vector of ∂M at ϕ_0 with the vector $\phi_0 - \phi^{'}$ is equal to $\delta/(2R_{\phi_0}) + o(\delta)$, while the angle β'' between the tangent vector with the vector $\phi^{''} - \phi_0$ is $\delta/(R_{\phi_0}) + o(\delta)$, where R_{ϕ_0} is the radius of curvature of ∂M at ϕ_0. So for small δ, $\langle D_{\phi_0}(h(\phi^{'}, \phi_0) + h(\phi_0, \phi^{''})), \mathbf{t} \rangle = \cos\beta' - \cos\beta'' > 3\delta^2/(8R_{\max}^2) + O(\delta^3)$, with $R_{\max}$ the maximal radius of curvature of ∂M.

With the help of Proposition 3.5, we show the gradient vector of h_n on the boundary of $Q^{n,\epsilon}$ is directed inwards for sufficiently small ϵ, subsequently the existence of the two critical values and their extremality in Proposition 3.4 can be obtained by invoking the Lyusternik-Shnirel'man theory [21]. The boundary of $Q^{n,\epsilon}$ is characterised by the conversion of some of the inequalities in the ϵ-conditions into equalities. When $n = 2$, the boundary of $Q^{2,\epsilon}$ is given by $|\phi_2 - \phi_1|_{\partial M} = \epsilon$. Thus the derivative of h_2 along the inward normal direction $\mathbf{n} = (\mathbf{t}_1/\sqrt{2}, -\mathbf{t}_2/\sqrt{2})$ (the case $\mathbf{n} = (-\mathbf{t}_1/\sqrt{2}, \mathbf{t}_2/\sqrt{2})$ can be treated similarly) reads

$$\begin{aligned}
& \langle D_{\mathbf{n}} h_2(\phi_1, \phi_2), \mathbf{t}_1 \rangle + \langle D_{\mathbf{n}} h_2(\phi_1, \phi_2), \mathbf{t}_2 \rangle \\
= & \frac{1}{\sqrt{2}} \left(\langle D_{\phi_1} h_2(\phi_1, \phi_2), \mathbf{t}_1 \rangle - \langle D_{\phi_2} h_2(\phi_1, \phi_2), \mathbf{t}_2 \rangle \right) \\
= & \frac{1}{\sqrt{2}} \left(\langle D_{\phi_1} \left(h(O_{e_1}, \phi_1) + h(\phi_1, \phi_2) \right), \mathbf{t}_1 \rangle - \langle D_{\phi_2} \left(h(\phi_1, \phi_2) + h(\phi_2, O_{e_2}) \right), \mathbf{t}_2 \rangle \right),
\end{aligned}$$

where $\mathbf{t}_j$, $j = 1$ or 2, is the unit tangent vector at ϕ_j along ∂M. Let $\phi_0 = \phi_1$ and $\phi' = \phi_2$ in Proposition 3.5, then the angle between the tangent vector along ∂M at ϕ_1 with the vector $O_{e_1} - \phi_1$ has a positive lower bound, while β' is of order δ and can be made as small as we wish. Therefore, $\langle D_{\phi_1} h_2(\phi_1, \phi_2), \mathbf{t}_1 \rangle > 0$. Similarly, $\langle -D_{\phi_2} h_2(\phi_1, \phi_2), \mathbf{t}_2 \rangle > 0$. In sum, the gradient vector of h_2 points inwards. When $n \geq 3$, it follows exactly by the same proof as in [23] that the derivative of h_n along the inward normal $\mathbf{n}$ is positive if the computation does not involve term $h(O_{e_1}, \phi_1)$ or $h(\phi_n, O_{e_2})$, namely, if $D_{\mathbf{n}} h_n(\phi_1, \ldots, \phi_n) = D_{\mathbf{n}}(\sum_{j=1}^{n-1} h(\phi_j, \phi_{j+1}))$. If $D_{\mathbf{n}} h_n$ involves terms $h(O_{e_1}, \phi_1)$ or $h(\phi_n, O_{e_2})$, then use the same argument as the $n = 2$ case.

This completes the proof of Proposition 3.4. □

Corollary 3.6. *Let $n \geq 1$. There exists an open and dense subset of $(\mathcal{M}, \mathcal{O})$ such that if (M, O) is a billiard system in that subset and $\{O_{e_1}, O_{e_2}\} \subset O$, then point scatterers O_{e_1} and O_{e_2} can be connected by at least two geometrically distinct non-degenerate n-link basic AI-orbits. (O_{e_1} and O_{e_2} may be the same point.)*

Proof. Since the critical values of h_n on $Q^{n,\epsilon}$ cannot all be isolated maxima, we can slightly perturb (M, O) in $(\mathcal{M}, \mathcal{O})$ by simultaneously changing the shape of ∂M and the positions of $O_1, \ldots, O_K$ if necessary so that the two critical points obtained in Proposition 3.4 are non-degenerate and have different indices and give rise to n-link basic AI-orbits. Moreover, it is clear that there is an open subset containing (M, O) in $(\mathcal{M}, \mathcal{O})$ for which there exist at least two non-degenerate critical points of h_n with distinct indices and any of the two critical points gives rise to an n-link basic AI-orbit. □

Proof of Theorem 1.11.

(i) By Corollary 3.6, there are at least two non-degenerate n-link basic AI-orbits connecting O_1 to O_1 for any billiard system (M, O) lying in an open and dense subset of $(\mathcal{M}, \mathcal{O})$. By taking intersection of these open and dense subsets for n ranging from 1 to N, we obtain another open and dense subset $(\mathcal{M}, \mathcal{O})_N$ of $(\mathcal{M}, \mathcal{O})$ in which every (M, O) has at least two non-degenerate n-link basic AI-orbits connecting O_1 to O_1 for every $1 \leq n \leq N$. So, there are $2N$ number of geometrically distinct non-degenerate basic AI-orbits $\Gamma_1, \Gamma_2, \ldots, \Gamma_{2N}$ for O_1 for the system (M, O). A Γ_b, $1 \leq b \leq 2N$, can intersect perpendicularly with ∂U_1 at only one or two points (corresponding to

$\Gamma^{\circlearrowleft} = \Gamma^{\circlearrowright}$ or $\Gamma^{\circlearrowleft} \neq \Gamma^{\circlearrowright}$, respectively). Due to a symmetry of h_n, the number of perpendicular intersection of Γ_b with ∂U_1 must be two if Γ_b is of even-link. This is because if $(\phi_1, \phi_2, \ldots, \phi_n)$ is a critical point of the function h_n, so is $(\phi_n, \ldots, \phi_2, \phi_1)$. If the number of perpendicular intersection is one (like $\overline{O_2\Omega_5} \cdot \overline{\Omega_5\Omega_6} \cdot \overline{\Omega_6\Omega_5} \cdot \overline{\Omega_5 O_2}$ in Figure 8), it must be $\phi_1 = \phi_n$, $\phi_2 = \phi_{n-1}, \ldots, \phi_n = \phi_1$. In case n is even, then it must be $\phi_{n/2} = \phi_{n/2+1}$, but this violates the ϵ-conditions in Definition 3.3. (If $\phi_{n/2} = \phi_{n/2+1}$, then the even-link basic AI-orbit becomes odd-link, a contradiction.) Thus, all even-link basic AI-orbits come in pairs (corresponding to $\Gamma^{\circlearrowleft}$ and $\Gamma^{\circlearrowright}$). Since $\Gamma_1, \ldots, \Gamma_{2N}$ are all geometrically distinct, totally they perpendicularly intersect at least $3N$ points with the circle ∂U_1 when N is even, but $3N-1$ points when N is odd. In either even or odd N case, the least total number of perpendicular intersection is even. Now, there are two cases. If the actual number of perpendicular intersection is even, let the number be $2C$ for some positive integer C. If the actual number is odd, then delete one whose intersection is due to an odd-link basic AI-orbit of the kind $\Gamma^{\circlearrowleft} = \Gamma^{\circlearrowright}$, and consider the left even number (also assumed to be $2C$) of points. It is easy to see that $[\frac{3N}{2}] \leq C \leq 2N-1$. Let the $2C$ number of intersection points on ∂U_1 be labelled by $w_{-C}, w_{-C+1}, \ldots, w_{-1}$, $w_1, \ldots, w_{C-1}, w_C$ with a rule that if two points on ∂U_1 form an antipodal pair then they are labelled by w_{-j} and w_j for some $1 \leq j \leq C$. See Figure 5. By Definition 1.10, a

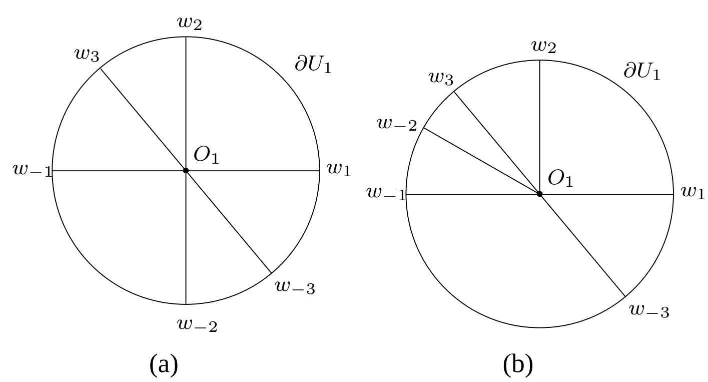

Figure 5: $N = 2$, $C = 3$. Points w_{-j} and w_j are not necessarily antipodal points.

product of $\cdots \Gamma_{b_i} \cdot \Gamma_{b_{i+1}} \cdots$ with $\Gamma_{b_i} \in \{\Gamma_1, \ldots, \Gamma_{2N}\}$ for every i is an N-link AI-orbit if $x_i^\dagger \in \{w_{-C}, \ldots, w_{-1}, w_1, \ldots, w_C\}$ and $y_i^\dagger \in \{w_{-C}, \ldots, w_{-1}, w_1, \ldots, w_C\}$ are not antipodal points of each other. Since $y_i^\dagger$ can be anything in $\{w_{-C}, \ldots, w_{-1}, w_1, \ldots, w_C\}$ except a particular one associated with $x_i^\dagger$, and since $y_i^\dagger$ will subsequently determine a unique $x_{i+1}^\dagger$, it follows that $x_{i+1}^\dagger$ can also be anything in $\{w_{-C}, \ldots, w_{-1}, w_1, \ldots, w_C\}$ except a particular one associated with $x_i^\dagger$. Therefore, there is a set of N-AI-orbits in $(\mathcal{M}, \mathcal{O})_N$ which corresponds bijectively to $\widetilde{\Sigma}_{2C}$.

(ii) Because the number of considered intersection points is finite, equal to $2C$, the least arc length of any two of these $2C$ points on ∂U_1 is at least some positive constant c. Therefore, we have the assertion (ii).

(iii) Since $z^\dagger = \{\ldots, (y_{i-1}^\dagger, x_i^\dagger), \ldots\}$ is uniquely determined by an AI-orbit, the inequality below holds

$$\max_{0 \le i < m} \left\{ |\tilde{y}^\dagger_{i-1} - y^\dagger_{i-1}|_{\partial U},\ |\tilde{x}^\dagger_i - x^\dagger_i|_{\partial U} \right\} > c$$

if and only if the two words $\{(\tilde{y}^\dagger_{-1}, \tilde{x}^\dagger_0), \ldots, (\tilde{y}^\dagger_{m-2}, \tilde{x}^\dagger_{m-1})\}$ and $\{(y^\dagger_{-1}, x^\dagger_0), \ldots, (y^\dagger_{m-2}, x^\dagger_{m-1})\}$ of length m are not identical. From the proof of assertion (i), we know that the set of sequences $z^\dagger$'s determined by N-AI-orbits is also bijective to $\widetilde{\Sigma}_{2C}$. Assertion (iii) therefore follows immediately. □

The above proof is already enough for Theorem 1.11. Nonetheless, we analyse one more case that all basic AI-orbits are connecting O_1 with O_2. Now there is an open and dense subset of $(\mathcal{M}, \mathcal{O})$ in which every billiard system possesses $2N+1$ number of non-degenerate basic AI-orbits $\Gamma_1, \Gamma_2, \ldots, \Gamma_{2N+1}$ connecting up O_1 with O_2 in which one is of 0-link, two are of 1-link, two are of 2-link, etc., and two are of N-link. These $2N+1$ basic AI-orbits intersect perpendicularly with ∂U_1 and ∂U_2 both at $2N+1$ points. Label those $2N+1$ points on ∂U_1 by $w_{-N}, \ldots, w_{-1}, w_0, w_1, \ldots, w_N$ and by $\bar{w}_{-N}, \ldots, \bar{w}_{-1}, \bar{w}_0, \bar{w}_1, \ldots, \bar{w}_N$ those $2N+1$ points on ∂U_2 with a rule that if two points form an antipodal pair on ∂U_1 (resp. ∂U_2) then they are labelled by w_{-j} and w_j (resp. $\bar{w}_{-j}$ and $\bar{w}_j$) for some $1 \le j \le N$, see Figure 6. Now, $\cdots \cdot \Gamma_{b_i} \cdot \Gamma_{b_{i+1}} \cdot \cdots$

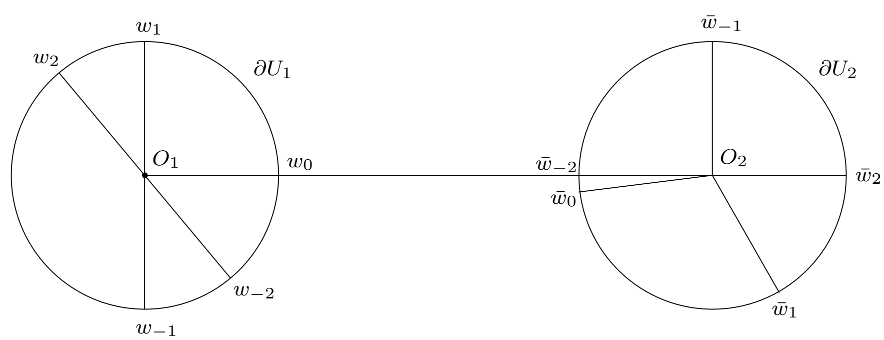

Figure 6: $N = 2$.

with $\Gamma_{b_i} \in \{\Gamma_1, \ldots, \Gamma_{2N+1}\}$ is an N-AI-orbit if $x^\dagger_i$ and $y^\dagger_i$ are not antipodal points. By our construction in Definition 1.10, $x^\dagger_i \in \{w_{-N}, \ldots, w_0, \ldots, w_N\}$ if and only if $y^\dagger_i \in \{w_{-N}, \ldots, w_0, \ldots, w_N\}$, also $x^\dagger_i \in \{\bar{w}_{-N}, \ldots, \bar{w}_0, \ldots, \bar{w}_N\}$ if and only if $y^\dagger_i \in \{\bar{w}_{-N}, \ldots, \bar{w}_0, \ldots, \bar{w}_N\}$. Without loss of generality, we may assume $x^\dagger_0 \in \partial U_1$ and subsequently $y^\dagger_0 \in \partial U_1$. By our construction again, $y^\dagger_0$ will uniquely determine the point $x^\dagger_1 \in \partial U_2$, and subsequently we have $y^\dagger_1 \in \partial U_2$. Since $y^\dagger_1$ will also uniquely determine $x^\dagger_2 \in \partial U_1$ and so on, we obtain a subshift of finite type with $4N+2$ symbols for $x_i^\dagger$ generated by the following diagram:

$$
\begin{array}{ll}
\{w_{-N},\dots,w_0,\dots,w_N\} \ni x_0^\dagger = w_{c_0} & : 2N+1 \text{ number of choices} \\
\qquad\qquad\qquad\downarrow & \\
\{w_{-N},\dots,w_0,\dots,w_N\} \ni y_0^\dagger \neq w_{-c_0} & : 2N \text{ number of choices} \\
\qquad\qquad\qquad\downarrow & \\
\{\bar{w}_{-N},\dots,\bar{w}_0,\dots,\bar{w}_N\} \ni x_1^\dagger = \bar{w}_{c_1} & : 1 \text{ choice (uniquely decided by } y_0^\dagger) \\
\qquad\qquad\qquad\downarrow & \\
\{\bar{w}_{-N},\dots,\bar{w}_0,\dots,\bar{w}_N\} \ni y_1^\dagger \neq \bar{w}_{-c_1} & : 2N \text{ number of choices} \\
\qquad\qquad\qquad\downarrow & \\
\{w_{-N},\dots,w_0,\dots,w_N\} \ni x_2^\dagger = w_{c_2} & : 1 \text{ choice (uniquely decided by } y_1^\dagger) \\
\qquad\qquad\qquad\downarrow & \\
\qquad\qquad\qquad\ddots &
\end{array}
$$

Therefore, the number of words of length m for $\{x_0^\dagger,\dots,x_{m-1}^\dagger\}$ in this subshift of finite type is $2(2N+1)(2N)^{m-1}$.

4 Dynamics near scatterers

For a given $1 \leq e \leq K$, we deal with in this section how the billiard orbits behave inside the domain U_e, which contains the scatterer A_e. Because the dynamics is local, we drop the subscript e throughout this section, and use the notations $O = O_e$, $\rho = \rho_e$. (Thus, O and ρ are scalars rather than vectors.) We assume B_ρ and U are discs of diameters, respectively, ρ and R centred at the origin O, and use x, y to indicate points on the boundary of U,

$$\partial U := \{y(\theta) = (R/2\cos\theta, R/2\sin\theta) : \ 0 \leq \theta < 2\pi\}, \tag{18}$$

and use Ψ for points on ∂B_ρ. Then, a point x in ∂U may be represented by $(R/2\cos\theta_x, R/2\sin\theta_x)$, likewise a point Ψ in ∂B_ρ means $\Psi = (\rho/2\cos\theta_\Psi, \rho/2\sin\theta_\Psi)$.

If there is no confusion with the definition of A_e ($1 \leq e \leq K$), we denote by A_ρ a C^3-family of strictly convex C^3-domains of $\mathbb{R}^2$ such that A_ρ is contained in B_ρ for every ρ and that $A_\rho = O$ when $\rho = 0$. (See Figure 7.)

A point y of the boundary ∂U is said to be 'accessible' by another point x lying in the boundary if the chord $\overline{xy}$ does not intersect with the scatterer A_ρ. Certainly, x is accessible to y if and only if y is accessible to x. In Figure 7, those points located in the least open arc between points $\hat{x}_a$ and $\hat{x}_c$ are not accessible by x in the case $A_\rho = B_\rho$. The point $\hat{x}$ indicates the antipode of x. Some planar geometry shows the following.

Proposition 4.1. *A point $y \in \partial U$ is accessible by another point $x \in \partial U$ if*

$$|\theta_y - \theta_x| < \pi - 2\sin^{-1}\frac{\rho}{R}.$$

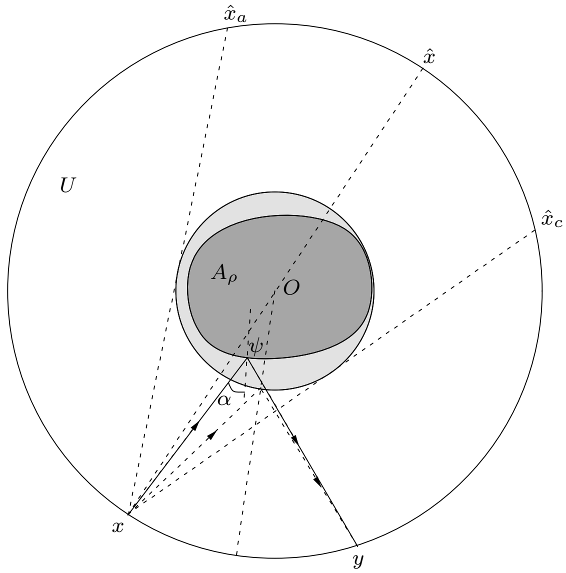

Figure 7: Local orbits inside U. Dark-shaded domain is A_ρ, while light-shaded domain is B_ρ.

Moreover, if x and y satisfy the above inequality when $\rho = \rho_0$ then they remain mutually accessible for $0 \le \rho < \rho_0$. In particular, all points are accessible to one another except their antipodes when $\rho = 0$.

Proposition 4.2. *Suppose points x and y in ∂U are accessible to each other if ρ is less than some constant ρ_0. Then there exists a unique local billiard orbit $\overrightarrow{x\psi} \cdot \overrightarrow{\psi y}$ connecting them with an only reflection point $\psi = \psi(x, y, \rho) \in \partial A_\rho$. Moreover, $h(x, \psi)$ and $h(\psi, y)$ depend C^2 on (x, y, ρ). In particular, $\lim_{\rho \to 0} h(x, \psi) = \lim_{\rho \to 0} h(\psi, y) = R/2$.*

Proof. ψ is equal to Ψ if A_ρ is B_ρ. From Figure 7, we know that $\theta_\Psi = (\theta_x + \theta_y)/2$, the angle $\angle xO\Psi$ is $(\theta_y - \theta_x)/2$, and that

$$h(x, \Psi) = h(\Psi, y) = \frac{1}{2}\sqrt{R^2 - 2R\rho\cos\frac{\theta_y - \theta_x}{2} + \rho^2}.$$

So the proposition follows. In the case $A_\rho \neq B_\rho$, we assume

$$\begin{aligned} \partial A_\rho := \{ q(\theta) = (g_\rho(\theta)\cos\theta, g_\rho(\theta)\sin\theta) : \\ 2\dot{g}_\rho^2 - g_\rho(\ddot{g}_\rho - g_\rho) > 0,\ 0 \le \theta < 2\pi \} \end{aligned} \tag{19}$$

with $g_\rho(\theta)$ depending jointly C^3 on θ and ρ. Condition (19) means ∂A_ρ has strictly positive curvature, or equivalently

$$\langle \ddot{q}(\theta), \mathbf{N}(q) \rangle > 0 \qquad \forall\, q \in \partial A_\rho, \tag{20}$$

where $\mathbf{N}(q)$ is the inward normal of ∂A_ρ at q. By the compactness of ∂A_ρ, the length function $h(x, q) + h(q, y)$ for fixed x and y with $q \in \partial A_\rho$ has a global minimum at some point, $\psi_{xx\rho}$ say, when $y = x$. In other words,

$$\langle D_q h(x,q) + D_q h(q,y), \dot{q} \rangle = 0 \tag{21}$$

at $q = \psi_{xx\rho}$. We have to show this minimum is unique and to show there is a unique family $\psi(x,\cdot;\rho)$ parametrised by ρ such that $q = \psi(x,y;\rho)$ solves (21) for fixed x. The second derivative of the length function reads

$$\langle (D^2_{qq} h(x,q) + D^2_{qq} h(q,y))\dot{q}, \dot{q} \rangle + \langle D_q h(x,q) + D_q h(q,y), \ddot{q} \rangle.$$

By Lemma 2.2, the first term is positive because it equals

$$|\dot{q}|^2 \frac{|q-x| + |q-y|}{|q-x|\,|q-y|} \cos^2 \alpha \tag{22}$$

and $\alpha = 0$ when $x = y$. (Here, α is the angle between $q - x$ and $\mathbf{N}(q)$. See Figure 7.) The second term is positive too, because of the convexity condition (20). Thereby, the implicit function theorem is applicable and such a function $\psi(x,y;\rho)$ exists for y near x, and ρ near ρ_0. Since (22) is positive for every $|\alpha| < \pi/2$ and for every $\rho < \rho_0$, this continuation can be carried on as long as the line segment $\overline{x\psi(x,y;\rho)}$ is not tangent to ∂A_ρ. □

5 Dynamics away from scatterers

We show in this section that $y^\dagger_{i-1}$ and $x^\dagger_i$ are not conjugate to each other if the AI-orbits defined in Definition 1.10 are non-degenerate. Therefore, any point near $y^\dagger_{i-1}$ can locally be connected to any point near $x^\dagger_i$ by a unique orbit segment. (See proposition 6.1.)

Proposition 5.1. *Suppose $\phi = \phi(\theta_\phi)$ and $y = y(\theta_y)$ are points respectively in ∂M and ∂U_e as described in (1) and (18) with O_e the centre of the polar coordinate system. Let the angle between $\phi - O_e$ and $\dot{\phi}$ be β (see Figure 3). Then*
(i)

$$\begin{aligned} D_y(h(O_e,y) + h(y,\phi)) &\equiv 0 \quad \textit{iff} \quad \theta_y = \theta_\phi; \\ \langle D_y(h(O_e,y) + h(y,\phi)), \dot{y} \rangle &= 0 \quad \textit{if} \quad \theta_y = \theta_\phi \pm \pi. \end{aligned} \tag{23}$$

(ii)

$$\langle D^2_{\phi y} h(y,\phi)\dot{\phi}, \dot{y} \rangle = -\langle D^2_{yy}(h(O_e,y) + h(y,\phi))\dot{y}, \dot{y} \rangle = -\frac{|\phi - O_e|\,|y - O_e|}{|\phi - y|}$$

at $\theta_y = \theta_\phi$.
(iii)

$$\sin\beta = \frac{|\phi - O_e|}{|\dot{\phi}|} \quad \textit{at } \theta_y = \theta_\phi.$$

Proof. (i) By Lemma 2.2, the vector

$$\frac{y-O_e}{|y-O_e|}-\frac{\phi-y}{|\phi-y|}$$

is identically zero if and only if O_e, $y(\theta_y)$ and $\phi(\theta_\phi)$ are collinear and $\theta_y = \theta_\phi$. If $\theta_y = \theta_\phi \pm \pi$, then the vector is perpendicular to the tangent vector $\dot{y}$.

(ii) The fact that (23) is true for every $\theta_y = \theta_\phi$ leads to

$$D^2_{yy}(h(O_e,y)+h(y,\phi))\dot{y}+D^2_{\phi y}(h(O_e,y)+h(y,\phi))\dot{\phi}=0$$

by taking derivative with respect to $\theta_y = \theta_\phi$. Therefore,

$$\begin{aligned}
&\langle D^2_{\phi y}h(y,\phi)\dot{\phi},\dot{y}\rangle = -\langle D^2_{yy}(h(y,O_e)+h(y,\phi))\dot{y},\dot{y}\rangle\\
&= \frac{\langle y-O_e,\dot{y}\rangle^2}{|y-O_e|^3}-\frac{|\dot{y}|^2}{|y-O_e|}+\frac{\langle \phi-y,\dot{y}\rangle^2}{|\phi-y|^3}-\frac{|\dot{y}|^2}{|\phi-y|} \quad \text{(by Lemma 2.2)}\\
&= -|\dot{y}|\frac{|\phi-O_e|}{|\phi-y|} \qquad \text{because} \quad \phi-O_e\perp\dot{y} \quad \text{and} \quad |y-O_e|\equiv|\dot{y}|.
\end{aligned}$$

(iii) Direct calculation via Lemma 2.2 yields

$$\begin{aligned}
&\langle D^2_{\phi y}h(y,\phi)\dot{\phi},\dot{y}\rangle\\
&\quad = \frac{-1}{|\phi-y|^3}\langle\phi-y,\dot{\phi}\rangle\langle\phi-y,\dot{y}\rangle-\frac{1}{|\phi-y|}\langle\dot{\phi},\dot{y}\rangle\\
&\quad = -\frac{|\dot{\phi}|\,|\dot{y}|}{|\phi-y|}\sin\beta \qquad \text{(see Figure 3)}.
\end{aligned}$$

Comparing with (ii), we obtain the assertion. □

Lemma 5.2. *The function $h_{n\oplus 2}:\partial U_{e_1}\times Q^{n,\epsilon}\times\partial U_{e_2}\to\mathbb{R}$, defined by*

$$h_{n\oplus 2}(y,\phi_1,\ldots,\phi_n,x):=h(O_{e_1},y)+h(y,\phi_1)+\sum_{j=1}^{n-1}h(\phi_j,\phi_{j+1})+h(\phi_n,x)+h(x,O_{e_2})$$

has a critical value at $(y,\Omega_1,\ldots,\Omega_n,x)$ with which $\overline{O_{e_1}y}\cup\overline{y\Omega_1}=\overline{O_{e_1}\Omega_1}$ and $\overline{\Omega_n x}\cup\overline{xO_{e_2}}=\overline{\Omega_n O_{e_2}}$ if and only if the function h_n defined in (17) has a critical value at $(\Omega_1,\ldots,\Omega_n)$ with which $\overline{O_{e_1}\Omega_1}\cap\partial U_{e_1}=y$ and $\overline{\Omega_n O_{e_2}}\cap\partial U_{e_2}=x$. The former is degenerate if and only if the latter is degenerate.

Proof. We remark first that

$$h_{1\oplus 2}(y,\phi_1,x):=h(O_{e_1},y)+h(y,\phi_1)+h(\phi_1,x)+h(x,O_{e_2})$$

and that

$$h_{0\oplus 2}(y,x):=h(O_{e_1},y)+h(y,x)+h(x,O_{e_2}) \qquad \text{(only when } O_{e_1}\neq O_{e_2}\text{)}.$$

The function $h_{n\oplus 2}$ has a critical value at point $(y,\Omega_1,\ldots,\Omega_n,x)$ if and only if

$$\langle D_y h_{n\oplus 2},\dot{y}\rangle=\langle D_{\phi_j}h_{n\oplus 2},\dot{\Omega}_j\rangle=\langle D_x h_{n\oplus 2},\dot{x}\rangle=0 \quad \forall j=1,\ldots,n \tag{24}$$

at that point. (Here $\dot{\Omega}_j$ stands for $\dot{\phi}(\theta)$ evaluated at $\phi = \Omega_j$. Similar notations are used for $\dot{y}$ and $\dot{x}$.) We know $D_y h_{n\oplus 2}(y, \Omega_1, \ldots, \Omega_n, x) = D_y(h(O_{e_1}, y) + h(y, \Omega_1))$, $D_x h_{n\oplus 2}(y, \Omega_1, \ldots, \Omega_n, x) = D_x(h(O_{e_2}, x) + h(x, \Omega_n))$. Therefore by Proposition 5.1, O_{e_1}, y and Ω_1 are collinear as well as O_{e_2}, x and Ω_n. The function h_n has a critical value at $(\Omega_1, \ldots, \Omega_n)$ if and only if

$$\langle D_{\phi_j} h_n(\Omega_1, \ldots, \Omega_n), \dot{\Omega}_j \rangle = 0 \quad \forall j = 1, \ldots, n. \tag{25}$$

Hence the first assertion of the lemma is true since both (24) and (25) imply that the law of specular reflection is complied with at $\Omega_1, \ldots, \Omega_n$.

By definition, $(y, \Omega_1, \ldots, \Omega_n, x)$ is a degenerate critical point if there exists non-zero $(\delta\theta_y, \delta\theta_1, \ldots, \delta\theta_n, \delta\theta_x) \in [0, 2\pi)^{n+2}$ such that

$$\begin{aligned} &\langle D^2_{yy}(h(O_{e_1}, y) + h(y, \Omega_1))\dot{y}\delta\theta_y + D^2_{\phi_1 y} h(y, \Omega_1)\dot{\Omega}_1\delta\theta_1, \dot{y}\rangle \\ &\quad + \langle D_y(h(O_{e_1}, y) + h(y, \Omega_1)), \ddot{y}\delta\theta_y\rangle = 0, \end{aligned} \tag{26}$$

$$\begin{aligned} &\langle D^2_{y\phi_1} h(y, \Omega_1)\dot{y}\delta\theta_y, \dot{\Omega}_1\rangle + \langle D^2_{\phi_1\phi_1}(h(y, \Omega_1) + h(\Omega_1, \Omega_2))\dot{\Omega}_1\delta\theta_1, \dot{\Omega}_1\rangle \\ &+ \langle D_{\phi_1}(h(y, \Omega_1) + h(\Omega_1, \Omega_2)), \ddot{\Omega}_1\delta\theta_1\rangle + \langle D^2_{\phi_2\phi_1} h(\Omega_1, \Omega_2)\dot{\Omega}_2\delta\theta_2, \dot{\Omega}_1\rangle \\ &= 0, \end{aligned} \tag{27}$$

$$\begin{aligned} &\langle \sum_{m=j-1}^{j+1} D^2_{\phi_m\phi_j}(h(\Omega_{j-1}, \Omega_j) + h(\Omega_j, \Omega_{j+1}))\dot{\Omega}_m\delta\theta_m, \dot{\Omega}_j\rangle \\ &\quad + \langle D_{\phi_j}(h(\Omega_{j-1}, \Omega_j) + h(\Omega_j, \Omega_{j+1})), \ddot{\Omega}_j\delta\theta_j\rangle = 0 \quad \forall j = 2, \ldots, n-1, \end{aligned} \tag{28}$$

$$\begin{aligned} & \langle D^2_{\phi_{n-1}\phi_n} h(\Omega_{n-1}, \Omega_n)\dot{\Omega}_{n-1}\delta\theta_{n-1}, \dot{\Omega}_n\rangle \\ + & \langle D^2_{\phi_n\phi_n}(h(x, \Omega_n) + h(\Omega_{n-1}, \Omega_n))\dot{\Omega}_n\delta\theta_n, \dot{\Omega}_n\rangle \\ + & \langle D_{\phi_n}(h(x, \Omega_n) + h(\Omega_{n-1}, \Omega_n)), \ddot{\Omega}_n\delta\theta_n\rangle + \langle D^2_{x\phi_n} h(x, \Omega_n)\dot{x}\delta\theta_x, \dot{\Omega}_n\rangle \\ = & 0, \end{aligned} \tag{29}$$

and

$$\begin{aligned} &\langle D^2_{xx}(h(O_{e_2}, x) + h(x, \Omega_n))\dot{x}\delta\theta_x + D^2_{\phi_n x} h(x, \Omega_n)\dot{\Omega}_n\delta\theta_n, \dot{x}\rangle \\ &\quad + \langle D_x(h(\Omega_n, x) + h(x, O_{e_2})), \ddot{x}\delta\theta_x\rangle = 0. \end{aligned} \tag{30}$$

Equations (26) and (30) together with (i) and (ii) of Proposition 5.1 lead to

$$\delta\theta_y = \delta\theta_1, \qquad \delta\theta_x = \delta\theta_n. \tag{31}$$

By definition, $(\Omega_1, \ldots, \Omega_n)$ is a degenerate critical point for h_n if the following identities are fulfilled

$$\begin{aligned} & \langle D^2_{\phi_1\phi_1}(h(O_{e_1}, \Omega_1) + h(\Omega_1, \Omega_2))\dot{\Omega}_1 d\theta_1, \dot{\Omega}_1\rangle \\ + & \langle D_{\phi_1}(h(O_{e_1}, \Omega_1) + h(\Omega_1, \Omega_2)), \ddot{\Omega}_1 d\theta_1\rangle \\ + & \langle D^2_{\phi_2\phi_1} h(\Omega_1, \Omega_2)\dot{\Omega}_2 d\theta_2, \dot{\Omega}_1\rangle = 0, \end{aligned} \tag{32}$$

$$\begin{aligned}&\langle \sum_{m=j-1}^{j+1} D^2_{\phi_m\phi_j}(h(\Omega_{j-1},\Omega_j)+h(\Omega_j,\Omega_{j+1}))\dot{\Omega}_m d\theta_m, \dot{\Omega}_j\rangle \\ &\quad +\langle D_{\phi_j}(h(\Omega_{j-1},\Omega_j)+h(\Omega_j,\Omega_{j+1})), \ddot{\Omega}_j d\theta_j\rangle = 0 \quad \forall j=2,\dots,n-1, \qquad (33)\end{aligned}$$

$$\begin{aligned}&\langle D^2_{\phi_{n-1}\phi_n} h(\Omega_{n-1},\Omega_n)\dot{\Omega}_{n-1} d\theta_{n-1}, \dot{\Omega}_n\rangle \\ + \quad &\langle D^2_{\phi_n\phi_n}(h(\Omega_{n-1},\Omega_n)+h(\Omega_n,O_{e_2}))\dot{\Omega}_n d\theta_n, \dot{\Omega}_n\rangle \\ + \quad &\langle D_{\phi_n}(h(\Omega_{n-1},\Omega_n)+h(\Omega_n,O_{e_2})), \ddot{\Omega}_n d\theta_n\rangle = 0 \qquad (34)\end{aligned}$$

for some non-zero $(d\theta_1,\dots,d\theta_n)$. Now take $(d\theta_1,\dots,d\theta_n)=(\delta\theta_1,\dots,\delta\theta_n)$. Subsequently, (28) is identical to (33). We know from Lemma 2.2 that

$$D_{\phi_1}h(y,\Omega_1)=\frac{\Omega_1-y}{|\Omega_1-y|}=\frac{\Omega_1-O_{e_1}}{|\Omega_1-O_{e_1}|}=D_{\phi_1}h(O_{e_1},\Omega_1),$$

thus identity (32) can be fulfilled if

$$\begin{aligned}&\langle D^2_{y\phi_1}h(y,\Omega_1)\dot{y},\dot{\Omega}_1\rangle+\langle D^2_{\phi_1\phi_1}h(y,\Omega_1)\dot{\Omega}_1,\dot{\Omega}_1\rangle \\ = \quad &\langle D^2_{\phi_1\phi_1}h(O_{e_1},\Omega_1)\dot{\Omega}_1,\dot{\Omega}_1\rangle \qquad (35)\end{aligned}$$

in (27). Direct calculation shows, with the help of Lemma 2.2, that

$$\begin{aligned}&\langle D^2_{\phi_1\phi_1}h(y,\Omega_1)\dot{\Omega}_1,\dot{\Omega}_1\rangle \\ &= \frac{-1}{|\Omega_1-y|^3}\langle\Omega_1-y,\dot{\Omega}_1\rangle^2+\frac{|\dot{\Omega}_1|^2}{|\Omega_1-y|}, \\ &= \frac{|\dot{\Omega}_1|^2}{|\Omega_1-y|}\sin^2\beta \quad \text{(see Figure 3)}, \qquad (36) \\ &\langle D^2_{\phi_1\phi_1}h(O_{e_1},\Omega_1)\dot{\Omega}_1,\dot{\Omega}_1\rangle \\ &= \frac{-1}{|\Omega_1-O_{e_1}|^3}\langle\Omega_1-O_{e_1},\dot{\Omega}_1\rangle^2+\frac{|\dot{\Omega}_1|^2}{|\Omega_1-O_{e_1}|}, \\ &= \frac{|\dot{\Omega}_1|^2}{|\Omega_1-O_{e_1}|}\sin^2\beta \quad \text{(see Figure 3)}. \qquad (37)\end{aligned}$$

Subtracting (36) from (37) and using Proposition 5.1, we arrive at

$$\begin{aligned}&\frac{-|y-O_{e_1}|}{|\Omega_1-y|\,|\Omega_1-O_{e_1}|}|\dot{\Omega}_1|^2\sin^2\beta \\ &= \frac{-|y-O_{e_1}|\,|\Omega_1-O_{e_1}|}{|\Omega_1-y|}=\langle D^2_{y\phi_1}h(y,\Omega_1)\dot{y},\ \dot{\Omega}_1\rangle.\end{aligned}$$

Thus, (35) is fulfilled. Likewise, we have

$$\begin{aligned}&\langle D^2_{x\phi_n}h(x,\Omega_n)\dot{x},\dot{\Omega}_n\rangle+\langle D^2_{\phi_n\phi_n}h(x,\Omega_n)\dot{\Omega}_n,\dot{\Omega}_n\rangle \\ = \quad &\langle D^2_{\phi_n\phi_n}h(O_{e_2},\Omega_n)\dot{\Omega}_n,\dot{\Omega}_n\rangle \qquad (38)\end{aligned}$$

in (29), and (34) is fulfilled thereby. Hence we have proved the “only if” case.

On the other hand, $h_{n\oplus 2}$ can be regarded as h_n if points O_{e_1}, y, Ω_1 are collinear with $\overline{O_{e_1}y}\cup\overline{y\Omega_1}=\overline{O_{e_1}\Omega_1}$ and O_{e_2} , x, Ω_n are collinear with $\overline{\Omega_n x}\cup\overline{xO_{e_2}}=\overline{\Omega_n O_{e_2}}$ too. Therefore if $(\Omega_1,\dots,\Omega_n)$ is a degenerate critical point of h_n, then $(y,\Omega_1,\dots,\Omega_n,x)$ with the mentioned collinearity constraints is a degenerate critical point of $h_{n\oplus 2}$. More precisely, suppose (32), (33) and (34) are satisfied for some non-zero $(d\theta_1,\dots,d\theta_n)=(\delta\theta_1,\dots,\delta\theta_n)$, then the collinearity constraints imply Proposition 5.1, (31), (35) and (38). Hence $h_{n\oplus 2}$ has a degenerate critical value at $(y(\theta_{\Omega_1}),\Omega_1,\dots,\Omega_n,x(\theta_{\Omega_n}))$ because (26)-(31) are fulfilled for $\delta\theta_y=\delta\theta_1=d\theta_1$, $\delta\theta_2=d\theta_2,\dots,\delta\theta_{n-1}=d\theta_{n-1}$, $\delta\theta_n=\delta\theta_x=d\theta_n$. □

6 Anti-integrability

In this final section, we prove Theorems 1.6 and 2.4.

Proposition 6.1. *Let $\{y^\dagger_{i-1},x^\dagger_i\}_{i\in\mathbb{Z}}$ be determined by a non-degenerate N-AI-orbit as in Definitions 1.10 and 3.2. There exist $\rho_0>0$ and neighbourhoods $\Delta_{y_{i-1}}\ni y^\dagger_{i-1}$, $\Delta_{x_i}\ni x^\dagger_i$ such that if $0\le\max_{1\le e\le K}\{\rho_e\}<\rho_0$ then for any two points y_{i-1}in $\Delta_{y_{i-1}}\subset\partial U_{i-1}$ and x_i in $\Delta_{x_i}\subset\partial U_i$, the function $h^*(y_{i-1},x_i)$ is C^2. Hence, the function F defined by (12) is C^1 on $Z\times[0,\rho_0)^K$ with*

$$Z:=\prod_{i\in\mathbb{Z}}\left(\Delta_{y_{i-1}}\times\Delta_{x_i}\right).$$

Proof. Suppose $(y^\dagger_{i-1},\Omega_1,\dots,\Omega_n,x^\dagger_i)$ is a non-degenerate critical point for $h_{n\oplus 2}$, then $(\Omega_1,\dots,\Omega_n)$ is a non-degenerate critical point for the function $h_{n\oplus 2}(y^\dagger_{i-1},\dots,x^\dagger_i)$ with fixed $y^\dagger_{i-1}$ and $x^\dagger_i$. Then, by the implicit function theorem, there are neighbourhoods $\Delta_{y_{i-1}}\ni y^\dagger_{i-1}$ and $\Delta_{x_i}\ni x^\dagger_i$, also there exists $\rho_0>0$ so that for $0\le\max_{1\le e\le K}\{\rho_e\}<\rho_0$ and for any pair of points (y_{i-1},x_i) in $\Delta_{y_{i-1}}\times\Delta_{x_i}$ we have that $(\Omega^*_1,\dots,\Omega^*_n)=(\omega_1(y_{i-1},x_i),\dots,\omega_n(y_{i-1},x_i))$ is the unique critical point for $h_{n\oplus 2}(y_{i-1},\dots,x_i)$ with fixed y_{i-1} and x_i for some C^2-functions $\omega_1,\dots,\omega_n$. Therefore, h^* is C^2. (Note that $\Delta_{y_{i-1}}$ and Δ_{x_i} can be chosen small enough so that $(y_{i-1},\Omega^*_1,\dots,\Omega^*_n,x_i)$ is a critical point for $h_{n\oplus 2}$ only when $y_{i-1}=y^\dagger_{i-1}$ and $x_i=x^\dagger_i$.)

From Proposition 4.2, we also know $h^-(x_{i-1},y_{i-1},\rho_{e_{i-1}})$ and $h^+(x_i,y_i,\rho_{e_i})$ are both C^2 dependent on their variables for every i. Hence, together with the prescribed number N, we have that $D_{z_i}h^-(x_{i-1},y_{i-1},\rho_{e_{i-1}})$, $D_{z_i}h^*(y_{i-1},x_i)$ and $D_{z_i}h^+(x_i,y_i,\rho_{e_i})$ all depend C^1 on their variables and uniformly with respect to i (because K and N are finite, and because each ∂U_e is fixed and compact). From the definition of F, we know $D_zF(z,\rho)=\{D_{z_j}F_i(z,\rho)\}_{i,j\in\mathbb{Z}}$ is a tri-diagonal infinite by infinite matrix, therefore F is C^1. □

Lemma 6.2. *Let $\{y^\dagger_{i-1},x^\dagger_i\}_{i\in\mathbb{Z}}$ be determined by an N-AI-orbit as in Definitions 1.10 and 3.2. The linear map $D_zF(z^\dagger,0):l_\infty\to l_\infty$ is invertible with bounded inverse if the N-AI-orbit is non-degenerate.*

Proof. Recall that zeros of the map $F(z,\rho)$ are governed by (13) and (14) or equivalently by (15) and (16). Now for all i, $h^-(x_i,y_i,0)$ and $h^+(x_i,y_i,0)$ are constant, equal to $R/2$, thus $D_zF(z,0)$ is diagonal, possessing diagonal terms $D_{z_i}F_i(z,0)$ equal to $D^2_{z_i}h^*(y_{i-1},x_i)$. Because $h(O_{e_{i-1}},y_{i-1}) = h(x_i,O_{e_i}) = R/2$ for all i, we have

$$D_{z_i}F_i(z,0) = D^2_{z_i}\left(h(O_{e_{i-1}},y_{i-1}) + h^*(y_{i-1},x_i) + h(x_i,O_{e_i})\right).$$

Note that, by our construction,

$$h^*(y_{i-1},x_i) = h(y_{i-1},\Omega^*_1) + \sum_{j=1}^{n-1} h(\Omega^*_j,\Omega^*_{j+1}) + h(\Omega^*_n,x_i),$$

where Ω^*_j, $j = 1,\dots,n$, are defined as those in the proof of Proposition 6.1. When $z = z^\dagger$ is determined by a non-degenerate N-AI-orbit, it turns out that $(\Omega^*_1,\dots,\Omega^*_n) = (\omega_1(y^\dagger_{i-1},x^\dagger_i),\dots,\omega_n(y^\dagger_{i-1},x^\dagger_i))$ is a non-degenerate critical point for $h_n(\phi_1,\dots,\phi_n) = h(O_{e_{i-1}},\phi_i) + \sum_{j=1}^{n-1} h(\phi_j,\phi_{j+1}) + h(\phi_n,O_{e_i})$ with prescribed $O_{e_{i-1}}$ and O_{e_i}. Thus $(y^\dagger_{i-1},\Omega^*_1,\dots,\Omega^*_n,x^\dagger_i)$ is a non-degenerate critical point of $h_{n\oplus 2}(y,\phi_1,\dots,\phi_n,x)$ by Lemma 5.2. Consequently, $D_{z_i}F_i(z^\dagger,0)$ is an invertible linear map from $\mathbb{R}^2\times\mathbb{R}^2$ to $\mathbb{R}^2\times\mathbb{R}^2$. Because this is true for every i, the map $D_zF(z^\dagger,0)$ is invertible. Since M and ∂M are compact and since the cardinality of the set O is the finite number K, the supremum of the norm of the inverse $D_{z_i}F_i(z^\dagger,0)^{-1}$ depends only on the number N and in particular it is independent of i. This means $D_zF(z^\dagger,0)^{-1}$ is bounded in the space l_∞. □

Proof of Theorem 2.4.

The theorem is a corollary of Proposition 6.1 and Lemma 6.2. □

Proof of Theorem 1.6.

Let N be an arbitrary positive integer. By Theorem 1.11, there is an open and dense subset of $(\mathcal{M},\mathcal{O})$ in which every billiard system (M,O) has a set $\mathfrak{X}$ of non-degenerate N-AI-orbits which corresponds bijectively to $\widetilde{\Sigma}_{2C}$ with $[\frac{3N}{2}] \le C \le 2N-1$. Every such a non-degenerate N-AI-orbit in (M,O) determines a unique sequence of pairs $z^\dagger = \{(y^\dagger_{i-1},x^\dagger_i)\}_{i\in\mathbb{Z}} \in \prod_{i\in\mathbb{Z}}(\partial U_{e_{i-1}}\times\partial U_{e_i})$ satisfying $F(z^\dagger,0) = 0$. Then Theorem 2.4 says that, by the implicit function theorem, there exist a unique $z^*(\rho)$ in a small neighbourhood of $z^\dagger$ in the space $\prod_{i\in\mathbb{Z}}(\partial U_{e_{i-1}}\times\partial U_{e_i})$ and a positive constant ρ_1, depending on (M,O), such that $F(z^*(\rho),\rho) = 0$ provided $0 \le \max_{1\le e\le K}\{\rho_e\} < \rho_1$. By Proposition 2.1, such a $z^*(\rho)$ corresponds a unique billiard orbit.

Let $\tilde{z}^\dagger$ be determined by another non-degenerate N-AI-orbit from $\mathfrak{X}$ satisfying $F(\tilde{z}^\dagger,0) = 0$, then there is a unique $\tilde{z}^*(\rho)$ in a small neighbourhood of $\tilde{z}^\dagger$ such that $F(\tilde{z}^*(\rho),\rho) = 0$ provided $0 \le \max_{1\le e\le K}\{\rho_e\} < \rho_1$. By Theorems 1.11 and 2.4, we know two facts. One fact is that $\tilde{z}^*(\rho) = \{\ldots, (\tilde{y}^*_{i-1}(\rho), \tilde{x}^*_i(\rho)), \ldots\}$ and $z^*(\rho) = \{\ldots, (y^*_{i-1}(\rho), x^*_i(\rho)), \ldots\}$ satisfy

$$\sup_{i\in\mathbb{Z}} \{|\tilde{x}^*_i(\rho) - x^*_i(\rho)|_{\partial U},\ |\tilde{y}^*_i(\rho) - y^*_i(\rho)|_{\partial U}\} > c_1$$

for some positive constant c_1. Let $\mathfrak{Y}_m$ be such a subset of the set consisting of the zeros of $F(\cdot, \rho)$ determined by applying the Implicit Function Theorem to all N-AI-orbits of $\mathfrak{X}$ that

$$\max_{0\leq i<m} \left\{|\tilde{y}^*_{i-1}(\rho) - y^*_{i-1}(\rho)|_{\partial U},\ |\tilde{x}^*_i(\rho) - x^*_i(\rho)|_{\partial U}\right\} > c_1 \tag{39}$$

for any pair of $\{\tilde{z}^*(\rho), z^*(\rho)\} \subset \mathfrak{Y}_m$. Then, the other fact is that the cardinality of $\mathfrak{Y}_m$ is exactly equal to $2C(2C-1)^{m-1}$.

Since a billiard orbit is uniquely determined by one of its historical state, every element $z^*(\rho)$ in $\mathfrak{Y}_m$ determines an element in $\partial(A_1 \cup \ldots \cup A_K) \times [-\pi/2, \pi/2]$ in an obvious way via Figure 7:

$$\{\ldots, (y^*_{i-1}(\rho), x^*_i(\rho)), \ldots\} \mapsto \{\ldots, (\psi^*_i, \alpha^*_i), \ldots\} \mapsto (\psi^*_{-1}, \alpha^*_{-1}),$$

where $\psi^*_i \in A_{e_i}$ if $x^*_i(\rho)$ and $y^*_i(\rho) \in \partial U_{e_i}$. Inequality (39) implies

$$\max_{-1\leq i<m} \{|\tilde{x}^*_i(\rho) - x^*_i(\rho)|_{\partial U},\ |\tilde{y}^*_i(\rho) - y^*_i(\rho)|_{\partial U}\} > c_1,$$

and this further implies

$$\max_{-1\leq i<m} \left\{|\widetilde{\psi}^*_i - \psi^*_i|_{\partial A},\ |\tilde{\alpha}^*_i - \alpha^*_i|\right\} > c_2$$

for some positive constant c_2, where $|\widetilde{\psi}^*_i - \psi^*_i|_{\partial A}$ stands for the least arc-length between $\widetilde{\psi}^*_i$ and ψ^*_i on ∂A_{e_i}. Therefore, for $0 \leq \max_{1\leq e\leq K}\{\rho_e\} < \rho_1$, the topological entropy of the first return map defined by (4) for the system (M, O) is at least $\ln(2C-1)$. $\square$

Proof of Remark 1.8.

The proof actually is a more detailed analysis of the proofs of Theorems 1.6 and 1.11, by taking the number of collisions occurred on the boundary ∂M into account.

As can readily be seen from the proof of Theorem 1.11 that the set $\mathfrak{X}$ of non-degenerate N-AI-orbits in the proof of Theorem 1.6 is comprised of $2N$ number of geometrically distinct basic AI-orbits $\{\Gamma_1, \ldots, \Gamma_{2N}\}$, in which two is of 1-link, two is of 2-link, ..., and two is of N-link. These $2N$ basic AI-orbits determine $2C$ number of points $\{(w_{-C}, 0), \ldots, (w_{-1}, 0), (w_1, 0), \ldots, (w_C, 0)\}$ on $\partial U_1 \times [-\pi/2, \pi/2]$, with $[\frac{3N}{2}] \leq C \leq 2N-1$. If we distinguish the dynamical difference between $\Gamma^{\circlearrowright}$ and $\Gamma^{\circlearrowleft}$, these $2N$ basic AI-orbits further give rise to P number of points on $\partial M \times [-\pi/2, \pi/2]$, with $3N^2/2 + N - 1/2 \leq P \leq 2N^2 + 2N - 2$ for $N \geq 2$. (Every n-link basic AI-orbit contributes n points $\phi_1, \phi_2, \ldots, \phi_n$ on ∂M with $(\phi_1, \ldots, \phi_n)$ a critical point of the length function h_n. Thus, $\Gamma^{\circlearrowright}$ together with $\Gamma^{\circlearrowleft}$ give $2n$ points if $\Gamma^{\circlearrowright} \neq \Gamma^{\circlearrowleft}$, but n points if $\Gamma^{\circlearrowright} = \Gamma^{\circlearrowleft}$. The bundle $[-\pi/2, \pi/2]$ corresponds to the incidence angles when $\Gamma^{\circlearrowright}$ or $\Gamma^{\circlearrowleft}$ collide with ∂M.) Because an N-AI-orbit has the form $\cdots \cdot \Gamma_{b_i} \cdot \Gamma_{b_{i+1}} \cdot \cdots$, with $\Gamma_{b_i} \in \{\Gamma_1, \ldots, \Gamma_{2N}\}$ for every integer i, the dynamical behaviour of $\mathfrak{X}$ on the $2C + P$ points on $(\partial U_1 \cup \partial M) \times [-\pi/2, \pi/2]$ is reminiscent of a topological Markov graph of $2C + P$ vertices in the following way: Assume $x_i^\dagger$ and $y_i^\dagger \in \{w_{-C}, \ldots, w_{-1}, w_1, \ldots, w_C\}$ are formed by $x_i^\dagger = \Gamma_{b_i} \pitchfork \partial U_1$ and $y_i^\dagger = \Gamma_{b_{i+1}} \pitchfork \partial U_1$ as described in Definition 1.10, then $y_i^\dagger \in \partial U_1$ can be anything in $\{w_{-C}, \ldots, w_{-1}, w_1, \ldots, w_C\}$ except a particular one associated with $x_i^\dagger$ (e.g. $y_i^\dagger \neq w_{-2}$ if $x_i^\dagger = w_2$), and along the basic AI-orbit $\Gamma_{b_{i+1}}$, $y_i^\dagger$ will subsequently determine n_{i+1} number, $n_{i+1} \leq N$, of points in ∂M (if $\Gamma_{b_{i+1}}$ is of n_{i+1}-link) and a unique point $x_{i+1}^\dagger$ in ∂U_1. In other words, we have the following topological Markov chain:

$$
\begin{array}{rl}
\partial U_1 \ni x_0^\dagger = w_{c_0} & : 2C \text{ choices} \\
\downarrow & \\
y_0^\dagger \neq w_{-c_0} & : 2C - 1 \text{ choices} \\
\downarrow & \\
\partial M \supset \Gamma_{b_1} \cap \partial M & : 1 \text{ choice } (n_1 \text{ points decided one by one along } \Gamma_{b_1}) \\
\downarrow & \\
\partial U_1 \ni x_1^\dagger = w_{c_1} & : 1 \text{ choice (uniquely decided by } y_0^\dagger = \Gamma_{b_1} \pitchfork \partial U_1) \\
\downarrow & \\
y_1^\dagger \neq w_{-c_1} & : 2C - 1 \text{ choices} \\
\downarrow & \\
\partial M \supset \Gamma_{b_2} \cap \partial M & : 1 \text{ choice } (n_2 \text{ points decided one by one along } \Gamma_{b_2}) \\
\downarrow & \\
\partial U_1 \ni x_2^\dagger = w_{c_2} & : 1 \text{ choice (uniquely decided by } y_1^\dagger = \Gamma_{b_2} \pitchfork \partial U_1) \\
\downarrow & \\
\vdots &
\end{array}
$$

The essence of the Markov chain is the following:

$$\cdots \xrightarrow{n_0+1 \text{ steps}} x_0^\dagger \xrightarrow{n_1+1 \text{ steps}} x_1^\dagger \xrightarrow{n_2+1 \text{ steps}} x_2^\dagger \longrightarrow \cdots .$$

Let $\widetilde{H}(m)$ denote the number of admissible words of length m in this $(2C + P)$-vertex topological Markov graph. We shall show that

$$\widetilde{H}(m) < 2C(2C-1)H(3m/2), \qquad \forall m \geq 2, \tag{40}$$

where $H(m)$ stands for the number of different broken lines in $\mathbb{R}^2$ which start at the origin, have length at most m, and have vertices at the integer points, such that the edges do not intersect any integer points except for the vertices, and that no three pairwise distinct consecutive vertices belong to one straight line, as described in [7]. Because

$\limsup_{m\to\infty} \ln H(m)/m$ is finite, as shown in [7], and because by (40)

$$\limsup_{m\to\infty} \frac{\ln \widetilde{H}(m)}{m} \le \limsup_{m\to\infty} \frac{\ln H(3m/2)}{m} = \frac{3}{2} \limsup_{m\to\infty} \frac{\ln H(m)}{m},$$

we conclude that the topological entropy of the $(2C+P)$-vertex Markov graph is finite, no matter how large C and P are. Then, following similar arguments in the proof of Theorem 1.6, we see that the billiard orbits obtained as the zeros of $F(\cdot,\rho)$ determined by applying the Implicit Function Theorem to all N-AI-orbits of $\mathfrak{X}$ do not lead to an arbitrarily large topological entropy of the billiard collision map (3), no matter how large N is.

In the sequel, we devote to (40). If we count multiplicity by distinguishing $\Gamma^{\circlearrowright}$ and $\Gamma^{\circlearrowleft}$, then the number of 1-link basic AI-orbits we are concerned with is 2, the number of n-link basic AI-orbits is 4 if n is even, and the number is 2, 3 or 4 if n is odd and at least three. Totally, we have $2C$ basic AI-orbits, with $[\frac{3N}{2}] \le C \le 2N-1$. Now, consider line segments that connect the origin $(0,0)$ with the integer points of coordinates $(n_i+1,\pm 1)$, $(-n_i-1,\pm 1)$, $(\pm 1, n_i+1)$, or $(\pm 1,-n_i-1)$ in $\mathbb{R}^2$ satisfying $1 \le n_i \le N$ for all $i \in \mathbb{Z}$. Next, consider the products $\cdots \cdot \overline{\mu_{-1}\mu_0} \cdot \overline{\mu_0\mu_1} \cdot \overline{\mu_1\mu_2} \cdot \cdots$ of these line segments and their translations, with a rule that $\mu_0 = (0,0)$, μ_i is an integer point in $\mathbb{R}^2$, $\overline{\mu_{i-1}\mu_i} \cdot \overline{\mu_i\mu_{i+1}}$ is not a straight line segment, and that μ_{i+1} has coordinates $(n_{i+1}+1,\pm 1)$, $(-n_{i+1}-1,\pm 1)$, $(\pm 1, n_{i+1}+1)$ or $(\pm 1, -n_{i+1}-1)$ if we translate the origin to μ_i. Subsequently, the number of admissible finite products $\overline{\mu_0\mu_1} \cdot \overline{\mu_1\mu_2} \cdot \cdots \cdot \overline{\mu_{l-1}\mu_l}$ is $8N(8N-1)^{l-1}$, which is greater than the number $2C(2C-1)^{l-1}$ of admissible finite products of basic AI-orbits $\Gamma_{b_1}\cdot\Gamma_{b_2}\cdot\cdots\cdot\Gamma_{b_l}$ (counting multiplicity $\Gamma^{\circlearrowright}$ and $\Gamma^{\circlearrowleft}$) that will form an N-AI-orbit if $l \to \infty$. Moreover, an easy observation shows two facts. The first is that any n-link basic AI-orbit consists of $n+1$ pieces of oriented line segments (or oriented edges); the second is that the length of every line segment $\overline{\mu_i\mu_{i+1}}$ is $\sqrt{n_{i+1}^2+2n_{i+1}+2}$, which is strictly between $n_{i+1}+1$ and $n_{i+1}+2$. These two facts imply that if an admissible finite product $\Gamma_{b_1}\cdot\Gamma_{b_2}\cdot\cdots\cdot\Gamma_{b_l}$ of basic AI-orbits is a product of m_1 number of consecutive oriented edges for some m_1 with $2l \le m_1 = n_1+1+n_2+1+\cdots+n_l+1 \le Nl+l$, then we can associate it with a unique admissible product $\overline{\mu_0\mu_1}\cdot\cdots\cdot\overline{\mu_{l-1}\mu_l}$ whose length is greater than m_1 but less than m_1+l. As a result, $\overline{H}(m_1) < 2CH(m_1+l) \le 2CH(3m_1/2)$, where $\overline{H}(m_1)$ stands for the number of admissible words of length m_1 starting from one of the $2C$ points $\{(w_{-C},0),\ldots,(w_{-1},0),(w_1,0),\ldots,(w_C,0)\}$ and ending at one of the same $2C$ points on $\partial U_1 \times [-\pi/2,\pi/2]$ in the aforementioned $(2C+P)$-vertex Markov chain. On the other hand, the number m of any admissible word of length $m \ge 2$ in the Markov chain must satisfy $m = \check{k}+n_1+1+n_2+1+\cdots+n_l+1+\hat{k} = \check{k}+m_1+\hat{k}$ for some $0 \le l \le m_1/2$, $0 \le m_1 \le m$, $0 \le \check{k} \le n_0 \le N$, and $0 \le \hat{k} \le n_{l+1} \le N$, for which there exists an admissible finite product $\Gamma_{b_0}\cdot\Gamma_{b_1}\cdot\cdots\cdot\Gamma_{b_l}\cdot\Gamma_{b_{l+1}}$ with Γ_{b_i} being of n_i-link for $0 \le i \le l+1$. As a result, $\widetilde{H}(m) \le (2C-1)\overline{H}(m) < 2C(2C-1)H(3m/2)$. $\square$

Remark 6.3. We finish this paper by demonstrating the capability of extending our study to non-convex billiard tables. Figure 8 shows a billiard table whose boundary curve consists of two Jordan curves ∂M_1 and ∂M_2. In the figure, we assume that Γ_1 is a 1-link basic AI-orbit for point-scatterer O_1, $\Gamma_2 = \Gamma_2^+$ or Γ_2^- is a 1-link basic AI-orbit connecting point-scatterers from O_1 to O_2 or from O_2 to O_1, while $\Gamma_3^{\circlearrowright}$ is a 2-link basic AI-orbit for O_2. We may assume all Γ_1, $\Gamma_2 = \Gamma_2^+$, $\Gamma_{-2} = \Gamma_2^-$, $\Gamma_3 = \Gamma_3^{\circlearrowright}$, and $\Gamma_4 = \Gamma_3^{\circlearrowleft}$ are non-degenerate basic AI-orbits by perturbing ∂M_1 if necessary. For any sequence $\{b_i\}_{i\in\mathbb{Z}}$ with b_i belonging to $\{1, \pm 2, 3, 4\}$ and with (b_i, b_{i+1}) not equal to $(1,-2)$, $(1,3)$, $(1,4)$, $(2,1)$, $(2,2)$, $(-2,-2)$, $(-2,3)$, $(-2,4)$, $(3,1)$, $(3,2)$, $(4,1)$, or $(4,2)$, we can construct an AI-orbit as the product of paths $\cdots \Gamma_{b_{i-1}} \cdot \Gamma_{b_i} \cdot \Gamma_{b_{i+1}} \cdots$. All such sequences $\{b_i\}_{i\in\mathbb{Z}}$ form a Markov chain of positive topological entropy. Arguing similarly to the proofs of Theorems 1.11 and 2.4, we can conclude that the system has positive topological entropy when O_1 and O_2 are fattened into small convex scatterers. Due to the non-convexity, $\overline{O_3\Omega_7} \cdot \overline{\Omega_7 O_3}$ is not a 1-link basic AI-orbit. Similarly, although the function $h_3 : (\phi_1, \phi_2, \phi_3) \mapsto h(O_2, \phi_1) + h(\phi_1, \phi_2) + h(\phi_2, \phi_3) + h(\phi_3, O_2)$ attains a critical value at $(\Omega_5, \Omega_6, \Omega_5)$, it dose not give rise to a 3-link basic AI-orbit. Nevertheless, if we remove M_2, then M_1 becomes a simply connected domain and is star-shaped with respect to O_2. In this situation, there are alway 1-link AI-orbits for O_2.

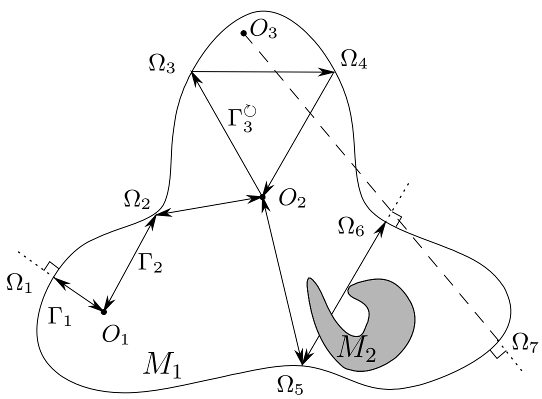

Figure 8: A billiard of non-convex, non-simply-connected domain with three point-scatterers.

Acknowledgments

The author is grateful to Serge Troubetzkoy for bringing the work [29] into his attention, to Anatole Katok and Mitsuhiro Shishikura for constructive comments. The author thanks Kyoto, Nanjing, and Suzhou universities for their hospitality during his visits.

References

[1] S. Aubry, G. Abramovici, Chaotic trajectories in the standard map: the concept of anti-integrability, Physica D **43** 199–219 (1990)

[2] S. Aubry, R.S. MacKay, C. Baesens, Equivalence of uniform hyperbolicity for symplectic twist maps and phonon gap for Frenkel-Kontorova models. Physica D **56** 123–134 (1992)

[3] G.D. Birkhoff, Dynamical systems. With an addendum by Jurgen Moser. Providence, R.I., American Mathematical Society Colloquium Publications, 1966

[4] A. Blokh, M. Misiurewicz, N. Simányi, Rotation sets of billiards with one obstacle. Commun. Math. Phys. **266**(1) 239–265 (2006)

[5] F.P. Boca, A. Zaharescu, The distribution of the free path lengths in the periodic two-dimensional Lorentz gas in the small-scatterer limit. Commun. Math. Phys. **269**(2), 425–471 (2007)

[6] L.A. Bunimovich, On ergodic properties of certain billiards. Funct. Anal. Appl. **8** 254–255 (1974)

[7] D. Burago, S. Ferleger, A. Kononenko, Topological entropy of semi-dispersing billiards. Ergod. Th. & Dynam. Sys. **18**(4) 791–805 (1998)

[8] Y.-C. Chen, Anti-integrability in scattering billiards. Dynamical Systems **19** 145–159 (2004)

[9] Y.-C. Chen, Bernoulli shift for second order recurrence relations near the anti-integrable limit. Discrete Contin. Dyn. Syst. B **5** 587–598 (2005)

[10] J. Cheng, Variational approach to homoclinic orbits in twist maps and an application to billiard systems. Z. Angew. Math. Phys. **55** 400–419 (2004)

[11] N.I. Chernov, Topological entropy and periodic points of two-dimensional hyperbolic billiards. Funct. Anal. Appl. **25** 39–45 (1991)

[12] N.I. Chernov, A new proof of Sinai's formula for computing the entropy of hyperbolic billiards. Its application to a Lorentz gas and stadium. Funct. Anal. Appl. **25** 204-219 (1991)

[13] P. Dahlqvist, The Lyapunov exponent in the Sinai billiard in the small scatterer limit. Nonlinearity **10** 159–173 (1997)

[14] V.J. Donnay, Elliptic islands in generalized Sinai billiards. Ergodic Theory Dynam. Systems **16**(5) 975–1010 (1996)

[15] C. Foltin, Billiards with Positive Topological Entropy. Bonn: Dissertation, Rheinischen Friedrich-Wilhelms-Universität Bonn, 2000

[16] C. Foltin, Billiards with positive topological entropy. Nonlinearity **15** 2053–2076 (2002)

[17] B. Friedman, Y. Oono, I. Kubo, Universal behavior of Sinai billiard systems in the small-scatterer limit. Phys. Rev. Lett. **52** 709–712 (1984)

[18] G. Gallavotti, D. Ornstein, Billiards and Bernoulli Schemes. Commun. Math. Phys. **38** 83–101 (1974)

[19] E. Gutkin, Blocking of billiard orbits and security for polygons and flat surfaces. Geom. Funct. Anal. **15**(1) 83–105 (2005)

[20] P. Hiemer, V. Snurnikov, Polygonal billiards with small obstacles. J. Stat. Phys. **90**(1–2) 453–466 (1998)

[21] H. Hofer, E. Zehnder, Symplectic Invariants and Hamiltonian Dynamics, Birkhauser Verlag, Boston, 1994

[22] A. Katok, B. Hasselblatt, Introduction to the Modern Theory of Dynamical Systems, Cambridge University Press, Cambridge, 1995

[23] V.V. Kozlov, D.V. Treshchev, Billiards. A Genetic Introduction to the Dynamics of Systems with Impacts, Translations of Mathematical Monographs, **89** American Mathematical Society, 1991

[24] V.F. Lazutkin, Existence of caustics for the billiard problem in a convex domain. Math. USSR Izv. **7** 185–214 (1973)

[25] M. Lenci, Typicality of recurrence for Lorentz gases. Ergod. Th. & Dynam. Sys. **26**(3) 799–820 (2006)

[26] A. Lopes, R. Markarian, Open billiards: invariant and conditionally invariant probabilities on Cantor sets. SIAM J. Appl. Math. **56** 651–680 (1996)

[27] R.S. MacKay, J.D. Meiss, Cantori for symplectic maps near the anti-integrable limit. Nonlinearity **5** 149–160 (1992)

[28] T. Monteil, A counter-example to the theorem of Hiemer and Snurnikov. J. Stat. Phys. **114**(5–6) 1619–1623 (2004)

[29] T. Morita, The symbolic representation of billiards without boundary condition. Trans. Amer. Math. Soc. **325** 819–828 (1991)

[30] A. Rapoport, V. Rom-Kedar, D. Turaev, Approximating multi-dimensional Hamiltonian flows by billiards. Commun. Math. Phys. **272**(3) 567–600 (2007)

[31] A. Rapoport, V. Rom-Kedar, D. Turaev, Stability in high dimensional steep repelling potentials. Commun. Math. Phys. **279**(2) 497–534 (2008)

[32] N. Simányi, Proof of the Boltzmann-Sinai ergodic hypothesis for typical hard disk systems. Invent. Math. **154**(1) 123–178 (2003)

[33] Ya G. Sinai, Dynamical systems with elastic reflections. Russ. Math. Surv. **25** 137-189 (1970)

[34] L. Stojanov, An estimate from above of the number of periodic orbits for semi-dispersed billiards. Commun. Math. Phys. **124** 217–227 (1989)

[35] L. Stoyanov, Exponential instability for a class of dispersing billiards. Ergod. Th. & Dynam. Sys. **19**(1) 201–226 (1999)

[36] P. Walters, An Introduction to Ergodic Theory, Springer-Verlag, New York, 1982