



An Upper Bound on Topological Entropy of the Bunimovich Stadium Billiard Map

Jernej Činč^{1,2} · Serge Troubetzkoy³

Received: 3 January 2023 / Accepted: 10 July 2023 / Published online: 25 August 2023
© The Author(s) 2023

Abstract

We show that the topological entropy of the billiard map in a Bunimovich stadium is at most $\log(3.49066)$.

Keywords Bunimovich stadium billiard · Topological entropy

Mathematics Subject Classification 37C83 · 37B40

1 Introduction

The Bunimovich stadium is a planar domain whose boundary consists of two semicircles joined by parallel segments as in Fig. 1. In this article we study the billiard in a Bunimovich stadium, this is the free motion of a point particle in the interior of the stadium with elastic collisions when the particle reaches the boundary. Billiards in stadia were first studied by Bunimovich in [6, 7] where he showed that the billiard has hyperbolic behavior and showed the ergodicity, K-mixing and Bernoulli property of the billiard map and flow with respect to the natural invariant measure (see also [14, 15]).

In this article we will study the topological entropy of the billiard map in a Bunimovich stadium. The topological entropy of a topological dynamical system is a real nonnegative number that is a measure of the complexity of the system. Roughly, it measures the exponential growth rate of the number of distinguishable orbits as time advances. We will discuss its exact definition in our setting in the next section.

Communicated by Marco Lenci.

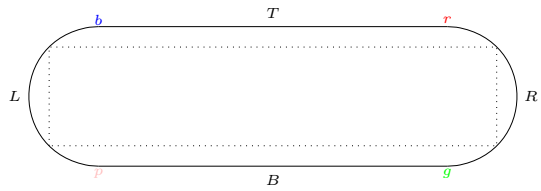
✉ Jernej Činč
jernej.cinc@um.si
Serge Troubetzkoy
serge.troubetzkoy@univ-amu.fr

¹ University of Maribor, Koroška 160, Maribor, Slovenia

² Centre of Excellence IT4Innovations - Institute for Research and Applications of Fuzzy Modeling, University of Ostrava, 30. Dubna 22, 701 03 Ostrava 1, Czech Republic

³ Aix Marseille Univ, CNRS, I2M, Marseille, France

Fig. 1 Labeling the sides of the stadium and a period 4 orbit



The study of topological entropy of billiards was initiated in [13]. In this article it was claimed with a one sentence proof that the topological entropy of the billiard map of stadia is at most $\log(4)$. A detailed proof using this strategy was given later by Bäker and Chernov, but they were able to show only a weaker estimate, that the topological entropy is at most $\log(6)$ [2]. Our main result will be a better upper bound on the topological entropy.

Recently, Misiurewicz and Zhang [18] have shown that as the side length tends to infinity the topological entropy of stadia is at least $\log(1 + \sqrt{2})$ by studying the map restricted to a subspace of the phase space which is compact and invariant under the billiard map. Another lower bound of the topological entropy can be derived from the variational principle¹ and the results of Chernov on the asymptotics of the metric entropy when the stadium degenerates to a circle, an infinite stadium, a segment, a point, or the plane in certain controlled ways [12].

Topological entropy of hyperbolic billiards has also been studied in several other articles [3, 8, 11, 21].

2 Definitions and Statement of the Results

We consider *the Bunimovich stadium billiard table* B_l , with the radius of the semicircles 1, and the lengths of straight segments $l > 0$. The *phase space* of this billiard map will be denoted by M_l . It consists of points s in the boundary of B_l and unit vectors pointing into the interior of B_l . We represent the unit vector by measuring its *angle* θ with respect to the inner pointing normal vector, thus

$$M_l := \{(s, \theta) : s \in \partial B_l, \theta \in (-\pi/2, \pi/2)\}.$$

The *billiard map* F_l is the first return map of the *billiard flow* Φ to the set M_l . Note that F_l is continuous, but M_l is not compact since we do not include vectors tangent to the boundary of B_l .

We remark that the map F_l does not extend to a continuous map of the closure of M_l . Thus all of the usual definitions of the topological entropy due to Adler, Konheim and McAndrew [1], Bowen [4, 5] and Dinaburg [17] can not be applied. There are several definitions of topological entropy which are possible. The definition we take, is a very natural one: we take a natural coding of the billiard, and then consider the entropy of the shift map on the closure of the set of all possible codes. This definition gives an upper bound of another natural definition of topological entropy on non-compact spaces, the Pesin-Pitskel' [20] topological entropy (this approach is closely related to that of Bowen given in [5], however Bowen's definition is not equivalent to the of Pesin-Pitskel', see [20][IV p. 308]). In particular, similar results for Sinai billiards (also known as Lorentz gas) were recently obtained by Baladi and Demers [3]. For a more detailed discussion of possible definitions of topological entropy in our setting and their relationship to our definition see Sect. 5.

¹ The variational principle holds for the Pesin-Pitskel' definition of entropy [20], see Sect. 5 for applicability to our situation.

We now give a precise definition of the topological entropy we consider. We label the four smooth components of the boundary by the alphabet $\{L, T, R, B\}$, the meeting points of the components have double labels (see Fig. 1). Slightly abusing notation we will say that $s \in Y$ where $Y \in \{L, T, R, B\}$ and mean that s is a point in ∂B_l with the label Y . It is easy to see that the corresponding partition is not a generator, for example the period 4 orbit with code $LLRR$ shown in Fig. 4 has the same code traced forward and backwards.

We consider two copies of L , denoted by \bar{L} and \underline{L} , (similarly \bar{R} and \underline{R} for R) and let $\bar{c} : M_l \rightarrow \mathcal{A} := \{\bar{L}, \underline{L}, T, B, \bar{R}, \underline{R}\}$ be the (multi-valued) coding map defined by $\bar{c}(s, \theta) = s$ if $s \in \{T, B\}$, $\bar{c}(s, \theta) = \bar{s}$ if $\theta \geq 0$ and $\bar{c}(s, \theta) = \underline{s}$ if $\theta \leq 0$ for $s \in \{L, R\}$. We consider the cover of the phase space into 6 elements given by this coding. The interiors of each element of the cover are disjoint, thus with the traditional misuse of terminology we will call this cover a *partition*.

We *code* the orbit of a point by the sequence of partition elements it hits, i.e.,

$$c(s, \theta) := (\omega_k)_{k \in \mathbb{Z}} \text{ where } \omega_k = \bar{c}(F_l^k(s, \theta)).$$

For $i \leq j$ let

$$\tilde{M}_l^{i,j} := \{(s, \theta) \in M_l : F_l^n(s, \theta) \text{ is in the interior of a partition element } \forall i \leq n \leq j\}.$$

Notice that since \bar{c} is multi-valued, the map c is multi-valued in particular on $\partial \tilde{M}_l^{i,j}$. However, for any point in the set $\tilde{M}_l^{i,j}$ the letter ω_k is unique for $i \leq k \leq j$, and thus for any point in the set

$$\tilde{M}_l := \bigcap_{i \leq j} \tilde{M}_l^{i,j}$$

the infinite coding is unique.

Let $\tilde{\Sigma}$ be the set of bi-infinite codes of points from \tilde{M}_l , and let $\bar{\Sigma}$ be the closure of $\tilde{\Sigma}$ in the product topology, and let $\mathcal{L}(n)$ be the set of words of length n appearing in $\tilde{\Sigma}$ (and thus in $\bar{\Sigma}$ as well). We let $p(n)$ denote the complexity of $\tilde{\Sigma}$; i.e.,

$$p(n) := \#\{(\omega_0, \dots, \omega_{n-1}) \in \mathcal{L}(n)\}.$$

The quantity $\log p(n)$ is sub-additive, thus the growth rate

$$\lim_{n \rightarrow \infty} \frac{\log p(n)}{n}$$

is well defined and is called the *topological entropy of the shift map restricted to the set $\bar{\Sigma}$* .

The 6 element partition is a *generating partition* in the sense that for each $\omega \in \bar{\Sigma} \setminus \{(TB)^\infty\}$ there is a unique $(s, \theta) \in M_l$ whose orbit has code ω (see [2] and the Appendix for a justification of this claim) thus it is natural to call this quantity the *topological entropy of the billiard map F_l* , i.e.,

$$h_{top}(F_l) := \lim_{n \rightarrow \infty} \frac{\log p(n)}{n}.$$

In this definition of the topological entropy we first miss a set by restricting to the interiors of partition elements, and then we add some points by taking the closure of $\tilde{\Sigma}$. The sequences in $\bar{\Sigma} \setminus \tilde{\Sigma}$ are all the codes of points which hit boundaries of partition elements obtained by using one sided continuity extension in the spatial coordinate. Although the entropy of $\bar{\Sigma}$ equals the entropy of $\tilde{\Sigma}$, we do not know anything about the Pesin-Pitsel' entropy of the invariant set $M_l \setminus \tilde{M}_l$ since the open (clopen) covers of $\bar{\Sigma}$ do not necessarily arise from an open cover of M_l . In particular we do not know if this entropy is smaller than the estimate from Theorem 1.

Our work was originally inspired by [18] where it was shown that

$$\lim_{l \rightarrow \infty} h_{top}(F_l) \geq \log(1 + \sqrt{2}) > \log(2.4142).$$

In fact in [18] the authors identify a certain compact subset of the phase space, such that if we restrict F_l to this set then we get equality in the above limit.

Another inspiration is [2]; the above mentioned fact about the six element partition being a generating partition immediately implies

$$h_{top}(F_l) \leq \log(6).$$

In the current paper we improve the upper bound on $h_{top}(F_l)$. Let $a := \frac{2W(\frac{1}{e})}{1+W(\frac{1}{e})}$ where $W(\frac{1}{e})$ is the unique solution to the equation $1 = we^{w+1}$, see [16] and the beginning of the proof of Lemma 8 for more information on the Lambert W function.

The main result of our article is the following theorem

Theorem 1 *For any $l > 0$ we have $h_{top}(F_l) < \log\left(2\left(\frac{2}{a} - 1\right)^a\right) < \log(3.49066)$.*

We prove Theorem 1 by studying possible word complexity of the 6 elements language associated to the Bunimovich billiard. In Sect. 3 we use Cassaigne’s formula from [9] and prove that $h_{top}(F_l)$ is bounded from above by the limit of logarithmic growth rate of the number of distinct saddle connections of increasing lengths. Cassaigne’s formula is very useful in studying low complexity systems, for example polygonal billiards [10]. To the best of our knowledge this is the first application of this formula to positive entropy systems. In Sect. 4 we give upper bounds for the number of different possible saddle connections using analytical tools, which yields our estimate for $h_{top}(F_l)$.

3 Saddle Connections

We consider the 6 element partition \mathcal{A} defined in the previous section. We will use the word *corner* to refer to the four points where the semi-circles meet the line segments as well as the two centers of the semi-circles. More formally, in the case of the centers of the semi-circles, by starting at a corner we mean that we start perpendicularly to a semi-circle and thus the flow passes through the corner when leaving the half-disk defined by the semi-circle. Recall that the four smooth components of the boundary of B_l are denoted by the alphabet $\{L, T, R, B\}$, see Fig. 1. The corners separate partition elements, this is clear for the four points, while for the centers of the semicircles we remark that the forward and backward orbit of any point $\bar{L} \cap \underline{L} = \{(s, \theta) : s \in L, \theta = 0\}$ passes through the center of the left semi-circle (a similar statement holds for points in $\bar{R} \cap \underline{R}$).

In analogy to polygonal billiards a *saddle connection* is an orbit segment which connects two corners of B_l (possibly the same) and does not visit any corner in between. To avoid technical complications, we do not consider the diameters of the semi-circles as saddle connections. The *length of a saddle connection* is the number of links in this trajectory. Except for saddle connections of length one we will represent a saddle connection by the first point of collision after leaving the starting corner, thus the code of the orbit segment of length $n - 1$ codes the saddle connection. Analogously, the empty word codes the saddle connections with length one. Let $N(n)$ denote the *number of distinct saddle connections of length at most n* and $\mathcal{N}(n)$ denote the *number of distinct saddle connections of length exactly n* . Our main result is based on the following result.

Proposition 2 $p(n) \leq 30 \sum_{j=0}^{n-1} N(j)$ for all $n \geq 1$.

Thus

$$h_{top}(F_I) = \lim_{n \rightarrow \infty} \frac{\log p(n)}{n} \leq \lim_{n \rightarrow \infty} \frac{\log (30 \sum_{j=0}^{n-1} N(j))}{n} = \lim_{n \rightarrow \infty} \frac{\log (\sum_{j=0}^{n-1} N(j))}{n},$$

which yields

Corollary 3

$$h_{top}(F_I) \leq \lim_{n \rightarrow \infty} \frac{\log (\sum_{j=0}^{n-1} N(j))}{n}.$$

To prove the proposition we need some techniques that were developed by Cassaigne in [9] and applied to polygonal billiards in [10]. Remember that $\mathcal{L}(n)$ is the set of blocks of length n in the subshift $\bar{\Sigma}$ (so $p(n) = \#\mathcal{L}(n)$). For $n \geq 1$, we define $s(n) := p(n + 1) - p(n)$. For $u \in \mathcal{L}(n)$ let

$$\begin{aligned} m_\ell(u) &:= \#\{a \in \mathcal{A} : au \in \mathcal{L}(n + 1)\}, \\ m_r(u) &:= \#\{b \in \mathcal{A} : ub \in \mathcal{L}(n + 1)\}, \\ m_b(u) &:= \#\{(a, b) \in \mathcal{A}^2 : aub \in \mathcal{L}(n + 2)\}. \end{aligned}$$

We remark that all three of these quantities are larger than or equal to one. A word $u \in \mathcal{L}(n)$ is called *left special* if $m_\ell(u) > 1$, *right special* if $m_r(u) > 1$ and *bispecial* if it is left and right special. Let

$$\mathcal{BL}(n) := \{u \in \mathcal{L}(n) : u \text{ is bispecial}\}.$$

In a more general setting in [9] (see [10] for an English version) it was shown that for all $k \geq 1$ we have

$$s(k + 1) - s(k) = \sum_{v \in \mathcal{BL}(k)} (m_b(v) - m_\ell(v) - m_r(v) + 1).$$

Consider the set of *strongly bispecial words*

$$\mathcal{BL}_s(n) := \{u \in \mathcal{L}(n) : u \text{ is bispecial and } m_b(v) - m_\ell(v) - m_r(v) + 1 > 0\}.$$

Clearly for all $k \geq 1$ we have

$$s(k + 1) - s(k) = \sum_{v \in \mathcal{BL}_s(k)} (m_b(v) - m_\ell(v) - m_r(v) + 1).$$

Proof of Proposition 2 Note that

$$(m_b(v) - m_\ell(v) - m_r(v) + 1) \leq \max_{0 \leq x, y \leq 6} (xy - x - y + 1) = 25,$$

thus summing over $1 \leq k \leq j - 1$ yields

$$s(j) \leq s(1) + 25 \sum_{k=1}^{j-1} \#\mathcal{BL}_s(k).$$

In our case $\#\mathcal{BL}_s(1) = p(1) = 6$ and $p(2) = 30$ since among the 36 possible words, the 6 words that can not be realized are $TT, BB, \bar{L}\bar{L}, \underline{L}\underline{L}, \bar{R}\bar{R}, \underline{R}\underline{R}$. Thus $s(1) = p(2) - p(1) = 30 - 6 = 24 = 4\#\mathcal{BL}_s(1)$, and thus we can estimate

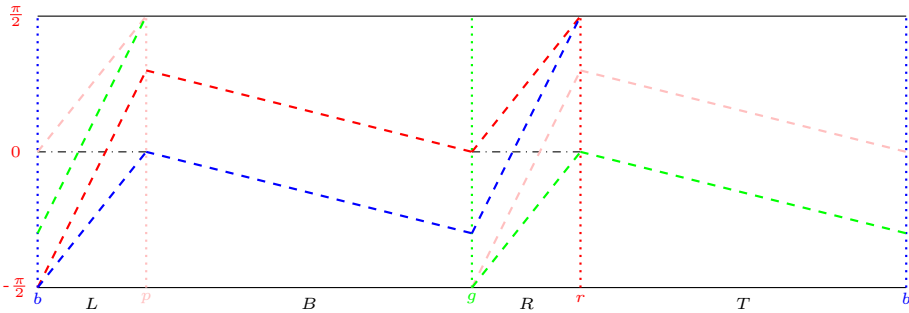


Fig. 2 The singularity sets of F_l (dotted) and F_l^{-1} (dashed). The singularity sets are monotone curves but for clarity they are drawn as linear segments. The two segments which are both dotted and dashed are in both singularity sets

$$s(j) \leq 29 \sum_{k=1}^{j-1} \#\mathcal{B}\mathcal{L}_s(k).$$

Remember that $s(j) = p(j + 1) - p(j)$, thus summing over $1 \leq j \leq n - 1$ yields

$$p(n) \leq p(1) + 29 \sum_{j=1}^{n-1} \sum_{k=1}^{j-1} \#\mathcal{B}\mathcal{L}_s(k).$$

Again we can adjust the constant to absorb the term $p(1)$ yielding the estimate

$$p(n) \leq 30 \sum_{j=1}^{n-1} \sum_{k=1}^{j-1} \#\mathcal{B}\mathcal{L}_s(k).$$

To finish the proof of Proposition 2 we need part a) from the following result. □

Proposition 4 a) For each $k \geq 1$ there is an injection $\mathcal{C} : \mathcal{B}\mathcal{L}_s(k - 1) \rightarrow \mathcal{N}(k)$.

b) For any $v \in \mathcal{B}\mathcal{L}(k)$ and any pair of corners (s, s') there is at most one saddle connection with code v starting at the corner s and ending at the corner s' .

Once (a) is proven, this yields

$$p(n) \leq 30 \sum_{j=1}^{n-1} \sum_{k=1}^{j-1} \#\mathcal{N}(k + 1) \leq 30 \sum_{j=0}^{n-1} N(j).$$

which completes the proof of Proposition 2.

Before proving Proposition 4 we need to introduce some more terminology. Let Γ' be the set of points from M_l perpendicular to the semicircles, i.e. $\Gamma' \subset M_l$ is the set of points for which $s \in L, R$ and $\theta = 0$. Let Γ be the union of Γ' with the set of points where F_l fails to be C^2 , and analogously Γ^- is the union of Γ' with the set of points where F_l^{-1} fails to be C^2 . We call Γ and Γ^- the singularity sets for F_l and F_l^{-1} respectively.

The singularity sets Γ, Γ^- consist of a finite number of C^1 -smooth compact curves in M_l (see Figure 3), which are increasing, decreasing, horizontal, or vertical. Define the singularity set for the map F_l^n for $n \geq 1$ by $\Gamma^n := \bigcup_{i=1}^n F_l^{-i+1}(\Gamma)$ and the singularity set for the map F_l^{-n} for $n \geq 1$ by $\Gamma^{-n} := \bigcup_{i=1}^n F_l^{i+1}(\Gamma^-)$.

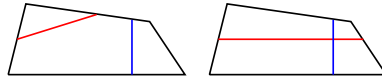


Fig. 3 Examples of $\omega(v)$ which are weakly and strongly bispecial with $m_\ell(v) = m_r(v) = 2$

Remember that the set \tilde{M}_l defined in Sect. 2 is the set of points in M_l having a well defined code, and \tilde{M}_l is the set of points whose orbit does not hit a corner.

For $(s, \theta) \in \tilde{M}_l$ let $c_k(s, \theta) := (\tilde{c}(F_l^i(s, \theta)))_{i=0}^{k-1}$ denote the *block of length k containing c(s, θ)*. For $v \in \mathcal{L}(k)$ we define the set

$$\omega(v) := \overline{\{(s, \theta) \in \tilde{M}_l : v = c_k(s, \theta)\}}$$

and call $\omega(v)$ a *k-cell*.

Proof of Proposition 4 (a) For $k = 1$ we note that the empty word is bispecial and it corresponds to 22 saddle connections, one for each pair of distinct corners excluding diameters and sides, and thus a) holds for $k = 1$.

Now suppose $k \geq 2$ and fix $v \in \mathcal{B}\mathcal{L}_s(k - 1)$. The proof of Lemma 2.5 in [2] shows that the set $\omega(v)$ is a simply connected closed set whose boundary consists of a finite collection of piecewise smooth curves with angles less than π at vertices. These curves belong to the union of the singular sets of F_l^i for $0 \leq i \leq k - 2$. For each $0 \leq i \leq k - 2$ the map F_l^i is continuous on $\omega(v)$.

Consider the “partition” $\bigcup_{avb \in \mathcal{L}(k+1)} F_l(\omega(avb))$ of the set $\omega(v)$, this is a partition in the sense that the interiors of the partition elements are pairwise disjoint. This partition is produced by cutting $\omega(v)$ by the singular sets of F_l^{k-1} and F_l^{-1} .

By assumption v is bispecial, so the branches of the singular set of F_l^{k-1} cut $\omega(v)$ into $m_r(v) \geq 2$ pieces and the branches of the singular set of F_l^{-1} cut $\omega(v)$ into $m_\ell(v) \geq 2$ pieces. Suppose first that these singularities do not intersect, then the union of these singular sets cut $\omega(v)$ into $m_r(v) + m_\ell(v) - 1$ pieces and thus the word v is weakly bispecial, i.e., $v \in \mathcal{B}\mathcal{L}(n) \setminus \mathcal{B}\mathcal{L}_s(n)$ and as mentioned above does not contribute to the sum, see Fig. 3 left.

Consider a point x of intersection of these two singular sets. As mentioned above the angle formed is less than π , i.e., the intersection must be transverse. If this intersection is on the boundary of a cell, then the orbit of x has at least three singular collisions, and so by definition x does not represent a saddle connection. So suppose x is in the interior of a cell $\omega(v)$. We have the preimage of x is a corner, and its forward image by F_l^{k-1} is also a corner, and all intermediate collisions are non-singular, thus it corresponds to a saddle connection of length $k - 1$. Thus for $k \geq 2$ we have verified part (a).

We turn to the verification of part (b). Label the corners of B_l by the alphabet $\mathcal{A}' := \{1, 2, 3, 4, 5, 6\}$. The *code of a saddle connection* is the sequence from $\mathcal{A} \cup \mathcal{A}'$ a point hits along with the starting and ending corners; thus a saddle connection of length n will have a code of length $n + 1$. To finish the proof we need to show that there is a bijection between saddle connections and their codes.

We will give a brief sketch describing the bijection between the set of codes and possible trajectories of the billiard map. A smooth curve from the phase space M_l equipped with a continuous family of unit normal vectors is called a *wave front*. Suppose by way of contradiction that two trajectories start at the same corner s' and end at the corner s'' (possibly $s'' = s'$) and have the same code. For concreteness the starting points are (s', θ_1) and (s', θ_2) . We consider the set $G := \{(s', \theta) : \theta_1 \leq \theta \leq \theta_2\}$ and for each $t \geq 0$ the corresponding wave front $G_t = \Phi_t(G)$. A wave front is said to *focus* at time $t > 0$ when the projection of the

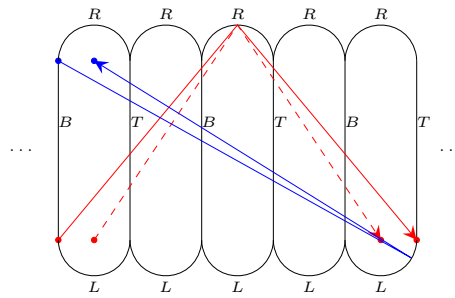


Fig. 4 Unfolding the stadium. Remember that centers of semi-circles are also corners. There are eight saddle connections with signed composition $2, 1, 2$, two of them having code $T\bar{B}RTB$ are drawn in red, there are two more saddle connections with this code, the other four have code $T\bar{B}LTB$. If we reverse the arrows we obtain the saddle connection with signed composition $-2, 1, -2$. In blue we show a saddle connection with code $T\bar{B}T\bar{B}L\bar{B}T\bar{B}T$ and signed composition $4, 1, -4$. If we reverse the arrows the blue saddle connection has the same code and signed composition

wave front G_t to the billiard table intersects itself. By way of contradiction we thus assumed the wave front G_t refocuses at a corner; we will show that this is in fact impossible. We refer to Subsection 8.4. in [14] for a more complete description of what follows.

Focusing occurs in Bunimovich billiards after the wave front reflects from one of the two semi-circles. Suppose that an infinitesimal wave front G_t collides with ∂M_l at some point in a semi-circle; denote the post-collisional curvature of the projection of G_t to the billiard table by \mathcal{G}_t^+ . The curvature of a wave front does not change at the instance of a collision with a flat boundary. Now suppose that the projection of G_t to the billiard table experiences collisions with semi-circles at times t and $t + \tau$, with possibly some flat collisions in between. Using (3.35) from [14] the wave front expands from a collision to another collision, if $|1 + \tau\mathcal{G}_t^+| > 1$. For this to hold, it is enough to check that $\mathcal{G}_t^+ < -2/\tau$ (see (8.2) from [14]). A focusing wave front with curvature $\mathcal{G}_t^+ < 0$ passes through a focusing point and defocuses at the time $t^* = t - 1/\mathcal{G}_t^+$ or in other words $\mathcal{G}_t^+ = \frac{1}{t-t^*}$ (see Section 3.8 in [14]). Thus $\mathcal{G}_t^+ < -2/\tau$ is equivalent to $t^* < t + \tau/2$. The last inequality indeed says that the wave front must defocus before it reaches the midpoint between the consecutive collisions. By Theorem 8.9. from [14] it holds that the families of unstable cones remain unstable under the iteration of the map. This implies that all wave fronts are in the unstable cones which gives a contradiction. Therefore, we indeed have unique coding of trajectories of the billiard map. \square

4 Proof of Theorem 1

To count the number of saddle connections we unfold the stadium (see Fig. 4). Consider an integer $j \geq 1$. We say $(n_1, m_1, n_2, m_2, \dots, n_k, m_k)$ is a *signed composition* of j if $n_i \in \mathbb{Z}$, $m_i \in \mathbb{N}$ with $m_i \geq 1$ for all i such that $\sum |n_i| + \sum m_i = j$. Let $Q(j)$ denote the number of *signed compositions* of j . Recall that $N(n)$ denotes the number of distinct saddle connections of length at most n .

Lemma 5 For each n and $l > 0$ we have

$$N(n) \leq 36 \sum_{j=0}^{n-1} Q(j).$$

Proof Fix a corner of B_l and consider the saddle connections of length at most n starting at this corner. We consider the associated signed composition in the following way: the non-negative integer $|n_i|$ counts the consecutive hits in the flat sides of B_l and m_i counts the consecutive hits in a semicircle. The sign of n_i tells us which way we are moving in the unfolding, left or right, when changing from one semicircle to the other. In this way each saddle connection yields a signed composition.

Fix $j \geq 1$ and a signed composition of j . As we showed in part b) from Proposition 4, for each pair of corners there is at most one saddle connection with this signed composition. Thus, since there are 6 corners, there are at most 36 codes of saddle connections which correspond to a given signed composition. \square

Let $(n_1, m_1, n_2, m_2, \dots, n_k, m_k)$ be a signed composition of j with $2k$ terms. Denote by $Q(j, k)$ the number of such possible compositions of $j \in \mathbb{N}$ with $2k$ terms. Let

$$r_i := |n_i| + m_i, \text{ then } \vec{r}_k(j) := (r_1, \dots, r_k)$$

is a composition of j with k terms. In what follows we will first estimate $Q(j, k)$ and then $Q(j)$.

Fix a composition $\vec{r}_k(j)$. Each $n_i \in \{-r_i + 1, \dots, -1, 0, 1, \dots, r_i - 1\}$ yields a different signed composition, there are

$$f(\vec{r}_k(j)) := \prod_{i=1}^k (2r_i - 1)$$

preimages of $\vec{r}_k(j)$ in total, i.e.,

$$Q(j, k) = \sum_{\ell \geq 1} \ell \times \#\{\vec{r}_k(j) : f(\vec{r}_k(j)) = \ell\}.$$

We start by estimating the number of terms in this sum, i.e., the largest possible value of ℓ . If $s = q_1 + \dots + q_k$, then the arithmetic–geometric mean inequality

$$\sqrt[k]{q_1 \cdots q_k} \leq \frac{q_1 + \dots + q_k}{k}$$

yields

$$q_1 \cdots q_k \leq \left(\frac{s}{k}\right)^k.$$

Notice that the equality is obtained if and only if all the $q_1 = q_2 = \dots = q_k$, and thus $q_i = s/k$. Setting $q_i = 2r_i - 1$ and $s = 2j - k$ yields

$$f(\vec{r}_k(j)) \leq \left(\frac{2j}{k} - 1\right)^k$$

with equality if and only if $r_i = \frac{2j}{k} - 1$.

Thus

$$\begin{aligned} Q(j, k) &\leq \left(\frac{2j}{k} - 1\right)^k \sum_{\ell \geq 1} \#\{\vec{r}_k(j) : f(\vec{r}_k(j)) = \ell\} \\ &= \left(\frac{2j}{k} - 1\right)^k \times \#\{\vec{r}_k(j)\} \\ &= \left(\frac{2j}{k} - 1\right)^k \binom{j}{k}. \end{aligned}$$

Fix $j \geq 1$ and let g_j be the function defined by

$$k \in \{1, \dots, j\} \mapsto \left(\frac{2j}{k} - 1\right)^k \binom{j}{k}.$$

Lemma 6 *The function $k \in \{1, \dots, j\} \mapsto \binom{j}{k}$ is increasing for $1 \leq k \leq \frac{j+1}{2}$ and decreasing for $\frac{j+1}{2} \leq k \leq j$.*

Proof The inequalities $1 \leq k - 1 < k \leq \frac{j+1}{2}$ imply that $\frac{j-k+1}{k} \geq 1$ and thus

$$\binom{j}{k} = \binom{j}{k-1} \frac{j-k+1}{k} \geq \binom{j}{k-1}$$

and thus the function is increasing for $1 \leq k \leq \frac{j+1}{2}$.

The decreasing statement holds since $\binom{j}{k} = \binom{j}{j-k}$. □

For each $1 \leq j$ let

$$h_j(x) = \left(\frac{2j}{x} - 1\right)^x = e^{x \ln\left(\frac{2j}{x} - 1\right)}.$$

Lemma 7 *For each $j \geq 2$ there exists a unique $x_j > 1$ such that h_j is increasing for $x \in [1, x_j]$ and decreasing for $x \in [x_j, j]$.*

Proof Throughout the proof the functions under consideration are restricted to the domain $[1, j]$. We begin by calculating the derivative of h_j ,

$$h'_j(x) = h_j(x) \left(\frac{-2j}{2j-x} + \ln\left(\frac{2j}{x} - 1\right)\right).$$

Let $k_j(x) := \frac{-2j}{2j-x} + \ln\left(\frac{2j}{x} - 1\right)$, then the signs of h'_j and k_j are the same since h_j is positive. We study the sign of k_j by taking its derivative:

$$k'_j(x) = -\frac{2j}{(2j-x)^2} - \frac{2j}{x(-x+2j)} = \frac{-4j^2}{x(2j-x)^2} < 0$$

for $x \in [1, j]$.

For $j \geq 2$ we have $k_j(1) = \frac{-2j}{2j-1} + \ln(2j-1) > 0$. Furthermore, $k_j(j) = -2 < 0$; thus there is a unique $x_j \in (1, j)$ which is the solution of the equation $k_j(x) = 0$ such that $\text{sgn}(h'_j(x)) = \text{sgn}(k_j(x)) > 0$ for $x \in (1, x_j)$ and $\text{sgn}(h'_j(x)) = \text{sgn}(k_j(x)) < 0$ for $x \in (x_j, j)$ and thus $h_j(x)$ is maximized when $x = x_j$. □

To prove the next lemma we will use the Lambert W function; see [16] for an introduction to this notion. The Lambert W function is a multivalued function which for a given complex number z gives all the complex numbers w which satisfy the equation $we^w = z$. If z is a positive real number then there is a single real solution w of this equation which we denote $W(z)$.

Lemma 8 *There exists a constant a such that $x_j = a \cdot j$ for each $j \geq 2$.*

Proof The equation $k_j(x_j) = 0$ is equivalent to $\frac{2j-x}{x} = e^1 e^{\left(\frac{x}{2j-x}\right)}$. Substituting $w = \frac{x}{2j-x}$ yields $1/w = e^1 e^w$ or equivalently $\frac{1}{e} = we^w$. Since $\frac{1}{e}$ is positive there is a single solution to this equation $w = W\left(\frac{1}{e}\right)$, and thus $x_j = \frac{2W\left(\frac{1}{e}\right)}{1+W\left(\frac{1}{e}\right)} \cdot j =: a \cdot j$. □

Lemma 9 *The constant a verifies $a \in (0.43562, 0.43563)$. The maximum value of h_j is at most $\left(\frac{2}{a} - 1\right)^{aj} < 1.74533^j$.*

Proof Notice that $k_j(0.43562j) \approx 0.00001 > 0$ and $k_j(0.43563j) \approx -0.00002 < 0$. Remembering from Lemma 7 that k_j is decreasing yields $0.43562 < a < 0.43563$. The maximum value of h_j is $h_j(x_j) = h_j(a \cdot j) = \left(\frac{2}{a} - 1\right)^{aj}$. If $0 < a_1 < a$ then $\frac{2}{a} - 1 < \frac{2}{a_1} - 1$. If furthermore $a < \min(a_2, 1)$ then $\frac{2}{a} - 1 > 1$ and thus $\left(\frac{2}{a} - 1\right)^a < \left(\frac{2}{a_1} - 1\right)^{a_2}$. Combining this with our previous estimate yields $\left(\frac{2}{a} - 1\right)^a < \left(\frac{2}{0.43562} - 1\right)^{0.43563} < 1.74533$. □

Combining Lemmas 6 and 9 yields:

Corollary 10 *For $j \geq 2$ the maximum value of function g_j is bounded from above by $\left(\frac{2}{a} - 1\right)^a \binom{j}{\lfloor \frac{j}{2} \rfloor} < 1.74533^j \binom{j}{\lfloor \frac{j}{2} \rfloor}$.*

To prove the next result we consider the gamma function $\Gamma(z)$, we will use the Legendre duplication formula

$$\Gamma(z)\Gamma\left(z + \frac{1}{2}\right) = 2^{1-2z}\sqrt{\pi}\Gamma(2z)$$

as well as Gautschi’s inequality

$$x^{1-s} < \frac{\Gamma(x+1)}{\Gamma(x+s)} < (x+1)^{1-s}$$

which holds for any positive real x and $s \in (0, 1)$.

Lemma 11 *For any even $j \geq 2$ we have*

$$\binom{j}{\lfloor \frac{j}{2} \rfloor} \leq \sqrt{\frac{2}{j}} \cdot \frac{2^j}{\sqrt{\pi}}$$

while for odd $j > 2$ we have

$$\binom{j}{\lfloor \frac{j}{2} \rfloor} \leq \sqrt{\frac{2}{j+1}} \cdot \frac{2^j}{\sqrt{\pi}}$$

Proof If $j = 2n$ is even then $\binom{j}{\lfloor \frac{j}{2} \rfloor} = \binom{2n}{n} = \frac{\Gamma(2n + 1)}{\Gamma(n + 1)^2}$. Using the duplication formula with $z = n + \frac{1}{2}$ yields

$$\frac{\Gamma(2n + 1)}{\Gamma(n + 1)^2} = \frac{\Gamma(2z)}{\Gamma(z + \frac{1}{2})^2} = \frac{\Gamma(z)}{\Gamma(z + \frac{1}{2})} \cdot \frac{2^{2z-1}}{\sqrt{\pi}} = \frac{\Gamma(\frac{j+1}{2})}{\Gamma(\frac{j}{2} + 1)} \cdot \frac{2^j}{\sqrt{\pi}}.$$

Next we apply Gautschi’s inequality with $s = \frac{1}{2}$ and $x = \frac{j}{2}$; it yields

$$\frac{\Gamma(\frac{j+1}{2})}{\Gamma(\frac{j}{2} + 1)} \cdot \frac{2^j}{\sqrt{\pi}} < \sqrt{\frac{2}{j}} \cdot \frac{2^j}{\sqrt{\pi}}.$$

Now suppose that $j = 2n + 1$ is odd, then

$$\binom{j}{\lfloor \frac{j}{2} \rfloor} = \binom{2n + 1}{n} = \frac{\Gamma(2n + 2)}{\Gamma(n + 1)\Gamma(n + 2)}.$$

Using the duplication formula with $z = n + 1$ yields

$$\frac{\Gamma(2n + 2)}{\Gamma(n + 1)\Gamma(n + 2)} = \frac{\Gamma(2z)}{\Gamma(z)\Gamma(z + 1)} = \frac{\Gamma(z + \frac{1}{2})}{\Gamma(z + 1)} \cdot \frac{2^{2z-1}}{\sqrt{\pi}} = \frac{\Gamma(\frac{j}{2} + 1)}{\Gamma(\frac{j+1}{2} + 1)} \cdot \frac{2^j}{\sqrt{\pi}}.$$

Again we apply Gautschi’s inequality, here with $x = \frac{j+1}{2}$ and $s = \frac{1}{2}$ which yields

$$\frac{\Gamma(\frac{j}{2} + 1)}{\Gamma(\frac{j+1}{2} + 1)} \cdot \frac{2^j}{\sqrt{\pi}} < \sqrt{\frac{2}{j + 1}} \cdot \frac{2^j}{\sqrt{\pi}}$$

□

Now we are ready to prove the main theorem of the paper.

Proof of Theorem 1 From Corollary 10 and Lemma 11 it follows that

$$Q(j, k) \leq \left(\left(\frac{2}{a} - 1 \right)^a \right)^j \binom{j}{\lfloor \frac{j}{2} \rfloor} \leq \left(\left(\frac{2}{a} - 1 \right)^a \right)^j \frac{2^j}{\sqrt{j\pi/2}} = \frac{\left(2 \left(\frac{2}{a} - 1 \right)^a \right)^j}{\sqrt{j\pi/2}}.$$

Therefore,

$$Q(j) \leq j \max_k(Q(j, k)) \leq \frac{\left(2 \left(\frac{2}{a} - 1 \right)^a \right)^j \sqrt{j}}{\sqrt{\pi/2}}.$$

Thus we obtain

$$N(n) \leq 36 \sum_{j=1}^{n-1} Q(j) \leq 36 \sum_{j=1}^{n-1} \frac{\left(2 \left(\frac{2}{a} - 1 \right)^a \right)^j \sqrt{j}}{\sqrt{\pi/2}} \leq \left(2 \left(\frac{2}{a} - 1 \right)^a \right)^n C \sqrt{n-1},$$

where C is a positive constant. Recall that $p(n)$ denotes the complexity of $\tilde{\Sigma}$. Therefore using Proposition 2,

$$p(n) \leq 30 \sum_{j=0}^{n-1} \left(2 \left(\frac{2}{a} - 1 \right)^a \right)^j C \sqrt{j-1} \leq \left(2 \left(\frac{2}{a} - 1 \right)^a \right)^n C' \sqrt{n-1}$$

where C' is another positive constant. From the definition of topological entropy we obtain

$$h_{top}(F_l) \leq \log \left(2 \left(\frac{2}{a} - 1 \right)^a \right) \leq \log(3.49066).$$

□

5 Other Possible Definitions of Topological Entropy

Another very natural definition of topological entropy was given by Pesin and Pitskel' in [20] and the closely related capacity topological entropy was defined by Pesin in [19][p. 75]. Applying these definitions to the map F_l restricted to \tilde{M}_l yields two quantities, the Pesin-Pitskel' topological entropy $h_{\tilde{M}_l}(F_l)$ and the capacity topological entropy $Ch_{\tilde{M}_l}(F_l)$. Formally, in [19] the capacity topological entropy is defined in a slightly more restrictive setting than in [20], but it can be defined in the setting of [20] and the relationship $h_{\tilde{M}_l}(F_l) \leq Ch_{\tilde{M}_l}(F_l)$ from [19] still holds. But our definition of $h_{top}(F_l)$ coincides with Pesin's definition of $Ch_{\tilde{M}_l}(F_l)$. Thus we have the following corollary

Corollary 12 *For any $l > 0$ the Pesin-Pitskel' topological entropy $h_{\tilde{M}_l}(F_l)$ is bounded from above by $\log \left(2 \left(\frac{2}{a} - 1 \right)^a \right) < \log(3.49066)$.*

Acknowledgements We thank Peter Bálint, Lyonia Bunimovich, Michał Misiurewicz and Yasha Pesin for useful discussions and helpful remarks. We also thank the referees for detailed constructive comments. J. Činč was partially supported by the FWF Schrödinger Fellowship stand-alone project J 4276-N35, the IDUB program no. 1484 "Excellence initiative - research university" for the AGH University of Science and Technology and the AD Program J1-4632 from ARRS, Slovenian National Research Agency.

Declarations

Conflicts of interest Data sharing is not applicable to this article as no datasets were generated or analysed during the current study. Both authors declare that they have no conflicts of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix

In [2] the authors state in Theorem 3.5 that a certain 16 element partition is generating. Just after the statement of the theorem they remark that it implies that the six element partition we consider in this article is a generating partition as well.

For convenience we give a formal proof of this remark. Consider the four element partition $\mathcal{B} := \{L, T, B, R\}$. The sixteen element partition \mathcal{C} consists of the connected components of the partition $\mathcal{C}' := \mathcal{B} \vee F_l^{-1}(\mathcal{B})$. The partition \mathcal{C}' has 14 elements since the codes TT and BB can not be realized. There are exactly two elements of \mathcal{C}' which are not connected,

corresponding to the pairs LL and RR . Each of them has two connected components, we call the splitting into the two connected components $\overline{LL}, \underline{LL}$, resp. $\overline{RR}, \underline{RR}$ which yields the partition \mathcal{C} . The component \overline{LL} (\overline{RR}) consists of $x \in M_I$ such that x and $F_I(x)$ are on the left (right) semi-circle of ∂B_I and $x = (s, \theta)$ with $\theta \leq 0$. Similarly, the component \underline{LL} (\underline{RR}) consist of such points where $\theta \geq 0$.

We consider the space of all codes $\tilde{\Sigma}_6$ resp. $\tilde{\Sigma}_{16}$ of orbits which stay in the interiors of the partition elements with the six letter, resp. 16 letter alphabet, and their closures $\overline{\Sigma}_6$ resp. $\overline{\Sigma}_{16}$. Note that $\tilde{\Sigma}_6$ resp. $\overline{\Sigma}_6$ is referred to $\tilde{\Sigma}$ resp. $\overline{\Sigma}$ in the rest of this article. Theorem 3.5 of [2] states that the partition \mathcal{C} is a generating partition in the following sense: there is a continuous surjection

$$\pi : \tilde{\Sigma}_{16} \setminus ((TB)^\infty \cup (BT)^\infty) \rightarrow \tilde{M}_I \setminus \hat{M}_I$$

such that

$$F_I \circ \pi = \pi \circ \sigma$$

where σ is the shift map and $\hat{M}_I := \{x \in M_I : c(s, \theta) = (BT)^\infty \text{ or } (TB)^\infty\}$.

Actually \mathcal{C} is a generating partition in a stronger sense; namely there is a continuous surjection

$$\pi : \overline{\Sigma}_{16} \setminus ((TB)^\infty \cup (BT)^\infty) \rightarrow M_I \setminus \hat{M}_I$$

such that $F_I \circ \pi = \pi \circ \sigma$. The proof of this stronger statement is identical to the proof of Theorem 3.5 of [2], and it is important to note that in Equation (10) of their proof the authors show that the intersection of closed cells is a single point.

We define $\phi : \overline{\Sigma}_6 \rightarrow \overline{\Sigma}_{16}$ by regrouping consecutive symbols and erasing the overbar or underbar except for the four cases $\overline{LL}, \underline{LL}, \overline{RR}, \underline{RR}$.

For example in Fig. 1 we consider the lower point on L pointing up, we have

$$\phi(\dots, \underline{L}, \underline{L}, \underline{R}, \underline{R}, \dots) = \dots, \underline{LL}, LR, \underline{RR}, RL, \dots$$

while if we consider the upper point on L pointing down, we have

$$\phi(\dots, \bar{L}, \bar{L}, \bar{R}, \bar{R}, \dots) = \dots, \overline{LL}, LR, \overline{RR}, RL, \dots$$

The map ϕ is continuous, surjective, and commutes with the shift map, thus

$$\pi \circ \phi : \overline{\Sigma}_6 \setminus ((TB)^\infty \cup (BT)^\infty) \rightarrow M_I \setminus \hat{M}_I$$

is a continuous surjection and such that $F_I \circ \pi \circ \phi = \pi \circ \phi \circ \sigma$, i.e., the six element partition is generating as well.

References

1. Adler, R.L., Konheim, A.G., McAndrew, M.H.: Topological entropy. *Trans. Am. Math. Soc.* **114**, 309–319 (1955)
2. Bäcker, A., Chernov, N.: Generating partitions for two-dimensional hyperbolic maps. *Nonlinearity* **11**, 79–87 (1998)
3. Baladi, V., Demers, M.: On the measure of maximal entropy for finite horizon Sinai billiard maps. *J. Am. Math. Soc.* **33**, 381–449 (2020)
4. Bowen, R.: Entropy for group endomorphisms and homogeneous spaces. *Trans. Am. Math. Soc.* **153**, 401–414 (1971)
5. Bowen, R.: Topological entropy for noncompact sets. *Trans. Am. Math. Soc.* **184**, 125–136 (1973)

6. Bunimovich, L.A.: The ergodic properties of certain billiards. *Funkcional. Anal. i Prilozhen* **8**, 73–74 (1974)
7. Bunimovich, L.A.: On the ergodic properties of nowhere dispersing billiards. *Commun. Math. Phys.* **65**, 295–312 (1979)
8. Burago, D., Ferleger, S., Kononeko, A.: Topological entropy of semi-dispersing billiards. *Ergodic Th. Dynam. Syst.* **18**, 791–805 (1997)
9. Cassaigne, J.: Complexité et facteurs spéciaux. *Bull. Belgian Math. Soc.* **4**, 67–88 (1997)
10. Cassaigne, J., Hubert, P., Troubetzkoy, S.: Complexity and growth for polygonal billiard. *Ann. Inst. Fourier (Grenoble)* **52**(3), 835–847 (2002)
11. Chen, Y.-C.: On topological entropy of billiard tables with small inner scatterers. *Adv. Math.* **224**, 432–460 (2010)
12. Chernov, N.: Entropy, Lyapunov exponents, and mean free path for billiards. *J. Stat. Phys.* **88**, 1–29 (1997)
13. Chernov, N.: Topological entropy and periodic points of two-dimensional hyperbolic billiards. *Funct. Anal. Appl.* **25**, 39–45 (1991)
14. Chernov, N., Markarian, R.: *Chaotic billiards*, Mathematical Surveys and Monographs, vol. 127, (2006)
15. Chernov, N., Troubetzkoy, S.: Ergodicity of billiards in polygons with pockets. *Nonlinearity* **11**, 1095–1102 (1998)
16. Corless, R.M., Gonnet, G.H., Hare, D.E.G., Jeffrey, D.J., Knuth, D.E.: On the Lambert W function. *Adv. Comput. Math.* **5**, 329–359 (1996)
17. Dinaburg, E.I.: A correlation between topological entropy and metric entropy. *Dokl. Akad. Nauk SSSR* **190**, 19–22 (1970)
18. Misiurewicz, M., Zhang, H.-K.: Topological entropy of Bunimovich stadium billiards. *Pure Appl. Funct. Anal.* **6**, 221–229 (2021)
19. Pesin, Y.B.: *Dimension theory in dynamical systems* Contemporary views and applications. Chicago Lectures in Mathematics. University of Chicago Press, Chicago, IL (1997)
20. Pesin, Ya..B., Pitskel, B.S.: Topological pressure and the variational principle for noncompact sets. *Funktsional. Anal. i Prilozhen* **18**, 50–63 (1984)
21. L. Stoyanov, *Instability and entropy for a class of dispersing billiards*, C.R. Acad. Sci. Paris Sér. I Math. **326** (1998) 737–742

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.