Uniqueness of billiard coding in polygons

Yunzhe Li\textsuperscript{a,b}

\textsuperscript{a}Department of Mathematics, Ecole Polytechnique, Palaiseau, France; \textsuperscript{b}Department of Mathematics, National University of Singapore, Singapore, Singapore

**ABSTRACT**

We consider polygonal billiards and we show that every nonperiodic billiard trajectory hits a unique sequence of sides if all the holes of the polygonal table have non-zero minimal diameters, generalizing a classical theorem of Galperin, Krüger and Troubetzkoy. Our approach uses symbolic dynamics and elementary geometry. We review some classical constructions in polygonal billiards and we introduce, as one of our main tools, a method to code pairs of parallel billiard trajectories in non-simply connected polygons. We also discuss some useful properties of ‘generalized trajectories’, which can be uniquely constructed from the limits of converging sequences of billiard codings.

**1. Introduction**

The study of mathematical billiards is a rich subject in dynamical systems. It describes the frictionless motion of a mass point in a domain with elastic reflection on the boundary. Among many other variations of mathematical billiards, polygonal billiards concerns the billiard problem in a two-dimensional polygonal domain.

Let \( Q \subset \mathbb{R}^2 \) be a polygon. The set \( TQ \) consists of couples \((x, v)\) where \( x \) is a point in \( Q \) and \( v \) a unit vector tangent to \( x \) representing the position and direction of the billiard ball, respectively. For a time \( t \geq 0 \) and an initial state \((x, v) \in TQ\) with \( v \) pointing strictly into \( Q \), the billiard flow \( \phi_t \) associates \((x, v)\) with a point \( \phi_t(x, v) \) in \( TQ \) obtained by moving \( x \) along a straight line in the direction of \( v \) with unit velocity until the moment \( t \). Whenever the trajectory enters the boundary \( \partial Q \) before time \( t \), the direction \( v \) is reflected by Descartes’ law (the angle of incident is equal to the angle of reflection). This motion is determined for all positive time \( t \) unless the flow reaches a vertex of the polygon.

One particular strategy to study polygonal billiards is to encode the trajectory of a billiard ball by the sequence of sides of the polygonal domain hit by the billiard ball along its trajectory. Let the set of edges of \( Q \) be labelled by a finite set \( A \). As illustrated in Figure 1, each infinite billiard trajectory \( \{\phi_t(x, v) \}_{t \geq 0} \) can be coded by a sequence \( \alpha \in A^\mathbb{N} \) labelling the edges that the trajectory hits in order. Conversely, if \( \alpha \in A^\mathbb{N} \) is a sequence, we define \( X(\alpha) \) to be the set of couples \((x, v) \in TQ\) with \( x \) lying on \( \partial Q \) such that the trajectory from \((x, v)\) is coded by \( \alpha \). This article aims to prove the following theorem.
Main Theorem: Let $Q$ be a polygon such that all the holes of $Q$ have nonzero minimal diameters. Suppose the edges of $Q$ are indexed by $A$. Then for any nonperiodic sequence $\alpha \in A^\mathbb{N}$, the set $X(\alpha)$ contains at most one point.

This result in the case of simply connected polygons was proven in [7]. This classical result has applications, for example, in [1, 2], where it was shown that the shape of a simply connected polygonal domain satisfying certain conditions could be uniquely determined, up to similarity, by the encoding of certain billiard trajectories. Other applications can be found in articles such as [3–5].

The proof in [7] relies on a particular property of simply connected polygon that $X(\alpha)$ consists of so-called ‘parallel phase points’ whose base points form a single interval on an edge. If $Q$ is not simply connected, however, the base points of $X(\alpha)$ may form more than one intervals due to the presence of ‘holes’ in $Q$ and consequently the method in [7] is not directly applicable to this case. Thus, we prove the main theorem for a more general class of polygons using a modified method. The main new idea in our approach is a way of coding parallel billiard trajectories (the $B$-codings introduced in Section 5) which captures certain information about the positions of holes relative to the trajectories being coded.

This article is structured as follows. In Section 2, we state the basic definitions of the dynamical system associated with polygonal billiards, and recall the technique of ‘trajectory unfolding’ – a useful visualization tool for later proofs. From Section 3 to Section 5, we set up various tools for the proof of the main theorem. In particular, Section 3 reviews some classical methods to code a billiard trajectory and introduces a partition of the phase space adapted to nonsimply connected polygons. In Section 4, we define ‘generalized trajectories’ as the limits of physical billiard trajectories, and discuss some of their properties. In Section 5, we introduce a new coding method adapted to parallel trajectories in nonsimply connected polygons. Especially, we prove Lemma 5.6 – a key result enabling the construction of generalized trajectories using the limit of trajectory codings. Finally, we put everything together in Section 6 to prove the main theorem.
2. The billiard map and trajectory unfolding

A polygon is defined to be a bounded connected region in $\mathbb{R}^2$ bounded by finitely many straight line segments, each of which is called an edge. The endpoints of the edges are called vertices. The union of the edges of a polygon $Q$ is called the boundary of $Q$, denoted by $\partial Q$. In general, a polygon may not be simply connected as shown in Figure 1. In this case, the complement of the interior of $Q$ in $\mathbb{R}^2$ contains finitely many bounded connected components, each of which will be called a hole in $Q$. Note that under the assumption of the main theorem that all holes of $Q$ have nonzero minimal diameters, the holes may still have an empty interior. For example, the holes may be in the form of a broken line. However, the holes must not be reduced to a single segment, i.e. a ‘slit’ in $Q$.

Recall that the set $TQ = Q \times S^1$ consists of couples $(x, v)$ where $x$ is a point in $Q$ and $v$ a unit vector tangent to $x$. We will call $x$ the base point of $(x, v)$. The billiard flow $\phi_t$ can be defined on a suitable subset of $TQ$. We define the phase space $V$ as a subset of $TQ$ consisting of the couples $(x, v)$ such that $x$ lies on $\partial Q$ and $v$ points strictly into $Q$. For $(x, v) \in V$, define $f$ to be the first return map to $V$ under the billiard flow, i.e. $f(x, v) = \phi_{t_0}(x, v)$ where $t_0 > 0$ is the smallest value such that $\phi_{t_0}(x, v) \in V$. The map $f$ is called the billiard map.

Following [9], we introduce a method to visualize billiard trajectories by ‘unfolding’. Let $p = (x, v)$ be a phase point with an infinite orbit $\{f^n(p) \mid n \geq 0\}$ under $f$. Let $Q_0 = Q$. Supposing the base point of $f(p)$ lies in $e$, we obtain a polygon $Q_1$ by reflecting $Q_0$ about the edge $e$. Define $\gamma_1$ to be the image in $Q_1$ of the line segment in $Q$ joining the base points of $p$ and $f(p)$. Continuing this way, we obtain a sequence of polygons $Q_n$ where $Q_n$ is a reflection of $Q_{n-1}$ about the edge $e_n'$ containing the base point of $f^n(p)$. The line segment $\gamma_n$ is the image in $Q_n$ of the line segment in $Q$ joining the base points of $f^n(p)$ and $f^{n+1}(p)$.

By identifying the edges of reflection between $Q_n$ and $Q_{n+1}$, an ‘infinite corridor’ $Q^\infty = \bigcup_{i \geq 0} Q_i$ can be constructed with a natural Riemannian metric inherited from $Q$. We will call $Q^\infty$ the unfolding of the billiard trajectory from $p$. Figure 2 illustrates the unfolding of a polygon $Q$. The edge shared between $Q_n$ and $Q_{n+1}$ will be called a reflecting edge. The infinite corridor $Q^\infty$ may or may not be embedded in $\mathbb{R}^2$ depending on the convexity of $Q$. For $0 \leq n \leq m \leq \infty$, put $Q_n^m := \bigcup_{i=n}^m Q_i$. If $m < \infty$, then we will call $Q_n^m$ a finite corridor. Let $\Gamma_p$ be the curve in $Q^\infty$ joining the line segments $\gamma_0, \gamma_1, \gamma_2, \ldots$ end to end. According to the law of reflection, $\Gamma_p$ is the trajectory in $Q^\infty$ of a straight line flow which begins from the point $x$ in $Q_0$ in the direction $v$.

Let $\Gamma$ be any ray contained in $Q^\infty$ pointing in the direction $(a, b) \in \mathbb{R}^2$. For example, we can take $\Gamma = \Gamma_p$ and let the direction of $\Gamma_p$ be the direction of the billiard trajectory from $p$. Relative to the direction $(a, b)$ of $\Gamma$, there is a natural way to define the ‘left side’ and the ‘right side’ of $\Gamma$ in $Q^\infty$. The following terminology will be frequently used: for $d > 0$, the open left (resp., right) $d$-strip of $\Gamma$ is defined to be the set of points in $Q^\infty \setminus \Gamma$ which can be obtained by translating a point in $\Gamma$ along the direction $(-b, a)$ (resp., in the direction $(b, -a)$) by a distance $< d$. The closed left/right $d$-strip of $\Gamma$ is defined as the closure of the left/right $d$-strip of $\Gamma$ in $Q^\infty$.

3. Partitions of the phase space

We follow the methods in [8] to parametrize the phase space $V \subset TQ$. Fix a vertex $v_0$ of $Q$ and a counter-clockwise orientation on the boundary $\partial Q$. Let $L$ be the total length of the
Figure 2. Unfolding a billiard trajectory into $Q^\infty = Q_0 \cup Q_1 \cup Q_2 \cup \cdots$. The shaded region shows an open left $d$-strip of the unfolded trajectory $\Gamma_p$.

boundary of $Q$. Each point on $\partial Q$ can be uniquely assigned a spatial coordinate $x \in [0, L)$ according to the orientation on $\partial Q$ such that $v_0$ corresponds to $x = 0$. Each phase point in $V$ can be assigned a pair $(x, \theta)$ with $0 < \theta < \pi$ and $0 \leq x < L$. Here, the angular coordinate $\theta$ of the phase point is the angle between its direction and the positive orientation on $\partial Q$. This pair $(x, \theta)$ is uniquely defined unless $x$ is a vertex of $Q$, in which case there are exactly two ways to assign $\theta$: it could be the angle measured from either of the two edges having $x$ as an endpoint. Let $x \in e$ for some edge $e \subset \partial Q$. Since the phase points always point into the polygon, it is natural to define the left side (resp., right side) of $x$ to be the set of points in $e$ whose spatial coordinates are smaller than (resp., greater than) that of $x$.

Let $\{e_a \mid a \in \mathcal{A}\}$ be the set of edges of $Q$ indexed by the finite set $\mathcal{A}$. We define a family $\{E_a\}_{a \in \mathcal{A}}$ of disjoint open subsets in $V$ where $E_a$ consists of phase points $(x, v)$ with $x$ lying in the interior of the edge $e_a$. Then the set

$$\overset{\circ}{V} = \bigcup_{a \in \mathcal{A}} E_a$$

is the set of phase points that are not based on a vertex and therefore have unique coordinates in the form $(x, \theta) \in [0, L) \times (0, \pi)$.

**Definition 3.1:** Let $p$ be a phase point such that $f^n(p)$ is defined up to $N \leq \infty$. The *edge coding of $p$* is the sequence $(\alpha_n)_{0 \leq n \leq N} \in \mathcal{A}^{N+1}$ satisfying

$$f^n(p) \in E_{\alpha_n}$$

for all $n = 0, 1, \ldots, N$.

The partition of $V$ into the disjoint union $\bigcup_{a \in \mathcal{A}} E_a$ does not provide geometric information about the presence of holes in a nonsimply connected domain. To remedy this, we
Figure 3. Both phase points $u$ and $v$ belong to $V_{a,b}$ because they are mapped from edge $e_a$ to edge $e_b$. They belong to two different connected components of $V_{a,b}$ due to the hole between their trajectories.

refine this partition by partitioning each $E_a$ further: for $a, b \in \mathcal{A}$, put

$$V_{a,b} = E_a \cap f^{-1}(E_b).$$

In other words, the subset $V_{a,b}$ consists of the phase points which will be sent from edge $e_a$ to edge $e_b$ by the billiard map $f$. If $Q$ is not simply connected, then $V_{a,b}$ may not be connected in $V$ as illustrated in Figure 3. Let

$$\{V^i_{a,b}\}_{i \in I_{a,b}}$$

be the collection of connected components of $V_{a,b}$ for a set $I_{a,b}$ depending only on $a, b$. The set $I_{a,b}$ is finite as there are finitely many holes in $Q$ that separate $V_{a,b}$. It follows that

$$\bigcup_{a, b \in \mathcal{A}} \bigcup_{i \in I_{a,b}} V^i_{a,b}$$

is a partition of $V$ into finitely many open subsets $V^i_{a,b}$ indexed by $\bigcup_{a, b \in \mathcal{A}} I_{a,b}$.

**Definition 3.2:** Let $p$ be a phase point such that $f^n(p)$ is defined up to $N \leq \infty$. Then $p$ can be uniquely associated with a sequence $\xi = (\xi_n)_{0 \leq n \leq N-1} \in (\bigcup_{a, b \in \mathcal{A}} I_{a,b})^N$ such that for $0 \leq n \leq N - 1$

$$f^n(p) \in V^i_{a,b} \text{ if and only if } \xi_n = i \in I_{a,b}.$$ 

We will call $\xi$ the $V$-coding of $p$.

We can verify the following result using the intermediate value theorem.

**Remark 3.3:** Let $a, b \in \mathcal{A}$ and suppose $p \in V^i_{a,b}$ and $q \in V^j_{a,b}$ are two phase points. Then $i = j$ if and only if there are no holes between the billiard trajectory from $p$ to $f(p)$ and the billiard trajectory from $q$ to $f(q)$. Hence, if two phase points have the same $V$-coding, then there are no holes between the trajectories from the two phase points.
4. Generalized trajectories and their codings

A billiard trajectory does not contain any vertex of the polygon. When the trajectory is unfolded into an infinite corridor (see Section 2), it intersects the reflecting edges transversally and never intersects nonreflecting edges. These properties are not necessarily preserved when we take the ‘limit’ of a sequence of billiard trajectories. In this section, we define the generalized trajectories which may not be physically realizable but may be thought of as the limit of physical trajectories.

Let \( f \) denote the billiard map on the polygon \( Q \). We have defined the partition

\[
\mathcal{V} = \bigcup_{a,b \in \mathcal{A}} \bigcup_{i \in I_{a,b}} V_{a,b}^i
\]

in Section 3. For \( a, b \in \mathcal{A} \) and \( i \in I_{a,b} \), let \( \overline{V_{a,b}^i} \) be the closure of \( V_{a,b}^i \) in \( TQ \). We call each element of the disjoint union

\[
\bigcup_{a,b \in \mathcal{A}} \overline{V_{a,b}^i}
\]

a generalized phase point.

**Definition 4.1:** A sequence of generalized phase points \( \{z_n\}_{n \geq 0} \) is said to be a generalized trajectory if for all \( n \geq 0 \) and \( a, b \in \mathcal{A} \) such that \( z_n \in \overline{V_{a,b}^i} \), there exists a sequence of phase points \( \{p_m\}_{m \geq 0} \subset V_{a,b}^i \) satisfying

\[
p_m \rightarrow z_n \quad \text{and} \quad f(p_m) \rightarrow z_{n+1} \quad \text{as} \quad m \rightarrow \infty.
\]

We call \( \{z_n\}_{n \geq 0} \) a physical trajectory if it is the orbit under \( f \) of some point in \( \mathcal{V} \).

We can extend the definitions of edge codings and \( V \)-codings to generalized trajectories in an obvious way. Using the ‘unfolding’ method in Section 2, we can similarly construct an infinite corridor \( Q^\infty \) according to the edge coding of a generalized trajectory \( \{z_n\}_{n \geq 0} \) and obtain an unfolded ‘trajectory’ \( \Gamma \) in \( Q^\infty \). The following properties of \( \Gamma \) can be easily deduced from Definition 4.1.

**Remark 4.2:** Let \( \{z_n\}_{n \geq 0} \) and \( \Gamma \) be defined as above. Then

1. For all \( n \geq 0 \), the interior of the segment in \( Q \) joining the base points of \( z_n \) and \( z_{n+1} \) does not transversally intersect the interior of any edge of \( Q \).
2. \( \Gamma \) is a straight line and \( \Gamma \) is contained in \( Q^\infty \).
3. If \( \Gamma \) does not contains any vertices, then \( \Gamma \) is a physical trajectory.

For convenience, we will also call the straight line \( \Gamma \) in \( Q^\infty \) associated with \( \{z_n\}_{n \geq 0} \) a generalized trajectory. Despite the nice properties of \( \Gamma \) in Remark 4.2, the generalized trajectory \( \Gamma \) may contain vertices in \( Q^\infty \) or even overlap one or more edges entirely as illustrated in Figure 4. Two particular classes of vertices lying on \( \Gamma \) are distinguished in the following definition.
**Figure 4.** The generalized trajectory $\Gamma$ contains $v_1$, $v_2$, $v_3$ and $v_4$. It overlaps the edge $v_2v_3$.

**Definition 4.3:** Let $\{z_n\}_{n \geq 0}$ be a generalized trajectory in $Q$ whose unfolded trajectory in $Q^\infty$ is $\Gamma$. Let $v$ be a vertex in $Q^\infty$ lying on $\Gamma$. We say $v$ blocks $\Gamma$ from the left (resp., from the right) if $v$ is an end vertex of an edge $e$ such that one of the followings holds

1. The edge $e$ is a reflecting edge, and the interior of $e$ lies on the right (resp., left) side of $\Gamma$ relative to the forward direction of $\Gamma$.
2. The edge $e$ is not a reflecting edge and the interior of $e$ lies on the left (resp., right) side of $\Gamma$ relative to the forward direction of $\Gamma$.

For example, the vertices $v_1v_2$ and $v_3$ in Figure 4 block $\Gamma$ from the right whereas $v_4$ blocks $\Gamma$ from the left. The vertices that block $\Gamma$ in the sense of Definition 4.3 ensure that trajectories near $\Gamma$ do not have the same $V$-coding as $\Gamma$. More precisely, we have the following lemma.

**Lemma 4.4:** Suppose a vertex $v \in Q_n$ in the unfolding $Q^\infty$ blocks $\Gamma$ from the left (resp., right). Then if $\Gamma'$ is another generalized trajectory whose $V$-coding coincides with the $V$-coding of $\Gamma$ up to at least the $(n + 1)$th term, then $v$ does not lie on the right side (resp., left side) of $\Gamma'$.

If $\Gamma$ does not contain any vertices blocking it from the left (resp., right) or overlap any reflecting edge, and there are no vertices in the open left (resp., right) $\epsilon$-strip of $\Gamma$, then any billiard trajectory in the open left (resp., right) $\epsilon$-strip of $\Gamma$ will have the same edge coding as $\Gamma$.

**Proof:** Let us assume that $v$ blocks $\Gamma$ from the left, the case where $v$ blocks $\Gamma$ from the right being similar. Let $\Gamma'$ be another generalized trajectory satisfying the condition in the first statement of the lemma. Suppose, on the contrary, that $v$ lies on the right side of $\Gamma'$. If $v$ is the end vertex of a nonreflecting edge $e$ in $Q_n$, then Definition 4.3 implies that the interior of $e$ lies on the left of $\Gamma$. By Remark 3.3, no holes of $Q_n$ lie between $\Gamma$ and $\Gamma'$. Thus, the edge $e$ must belong to the outer boundary of $Q$ (i.e. the unique connected component of $\partial Q$ not bounding a hole of $Q$) as illustrated in Figure 5(a). However, this would force $\Gamma'$ to intersect some nonreflecting edge on the outer boundary, contradicting the condition on the $V$-coding of $\Gamma'$. 

On the other hand, if \( v \) is an end vertex of a reflecting edge \( e \), then \( \Gamma \) and \( \Gamma' \) exit a copy of \( Q \) both through \( e \). By Definition 4.3, the other end vertex \( u \) of \( e \) must be on the right side of \( \Gamma \). The vertex \( u \) lies either on the trajectory \( \Gamma'' \) or on the left side of \( \Gamma' \). Geometrically, this means \( \Gamma'' \) and \( \Gamma \) hit \( e \) from different sides as illustrated in Figure 5(b). This contradicts the requirement that \( \Gamma \) and \( \Gamma' \) must exit a copy of \( Q \) both through \( e \). In conclusion, we deduce that \( v \) does not lie on the right side of \( \Gamma' \).

For the second statement, let \( l \) be a billiard flow in the open left \( \epsilon \)-strip of \( \Gamma \). We show that \( e \) is a reflecting edge in \( Q^\infty \) if and only if \( l \) intersects the interior of \( e \), thereby proving that \( l \) and \( \Gamma \) have the same edge coding. If \( e \) is a reflecting edge in \( Q^\infty \), then its end vertex \( w \) on the left side of \( \Gamma \) is not on \( \Gamma' \) for otherwise it would block \( \Gamma \) from the left or \( e \) would be contained in \( \Gamma' \), contradicting the assumption. Also, the vertex \( w \) is not in the left \( \epsilon \)-strip of \( \Gamma \) by assumption. Thus \( l \) also intersects \( e \).

Conversely, if \( l \) intersects the interior of an edge \( e \), then the assumption implies that a vertex \( v \) of \( e \) lies either on \( \Gamma \) or on the right side of \( \Gamma' \). In the latter case, \( \Gamma \) clearly intersect the interior of \( e \) and thus \( e \) is a reflecting edge. If \( v \) is on \( \Gamma \), then \( v \) must be a reflecting edge, for, if not, the edge \( e \) would block \( \Gamma' \) from the left by Definition 4.3, contradicting the assumption. We conclude that an edge \( e \) in \( Q^\infty \) is a reflecting edge if and only if \( l \) intersects the interior of \( e \), as desired. The proof of the analogous statement with ‘left’ replaced by ‘right’ is similar.

\[\square\]

5. Coding parallel trajectories

In this section, we introduce the main new tools in our proof: the coding of two parallel trajectories, which detects the positions of holes (Definition 5.1) and the construction of parallel generalised trajectories from a convergent sequence of codings (Lemma 5.6).

5.1. Definition of \( \mathcal{B} \)-codings

Before discussing parallel trajectories, let us consider a slightly more general situation. Suppose \( 0 \leq N \leq \infty \) and \( p_1 = (x_1, \theta_1) \) and \( p_2 = (x_2, \theta_2) \) are two phase points in \( V \) such
that the first $N + 1$ terms in their edge codings coincide. Then we can define a coding $(\beta_n)_{0 \leq n \leq N - 1}$ associated to the pair $(p_1, p_2)$ as follows. Let $(\xi_n)_{0 \leq n \leq N - 1}$ and $(\xi'_n)_{0 \leq n \leq N - 1}$ be the first $N$ terms of the $V$-codings associated with $p_1$ and $p_2$, respectively. If $f^n(p_1)$ is on the left side of $f^n(p_2)$ relative to the orientation on $\partial Q$, then we put $\beta_n = (\xi_n, \xi'_n)$. Otherwise, let $\beta_n = (\xi'_n, \xi_n)$. Thus defined, the sequence $\beta = (\beta_n)_{n \geq 0}$ is a sequence of elements in the finite set

$$B = \bigsqcup_{a,b \in A} I_{a,b} \times I_{a,b}.$$ 

**Definition 5.1:** The sequence $\beta$ constructed above is called the $B$-coding associated with the pair $(p_1, p_2)$.

Clearly, we can recover the edge coding and the $V$-codings of $p_1$ and $p_2$ from their $B$-coding $\beta$. Moreover, the coding $\beta$ also contains information about the holes between the two billiard trajectories from $p_1$ and $p_2$. The following lemma is a direct consequence of Remark 3.3 and the definition of $\beta$.

**Lemma 5.2:** Let $(\beta_n)_{0 \leq n \leq N - 1}$ be the $B$-coding associated with a pair of phase points $(p_1, p_2)$ as above. Suppose $\beta_n$ is given by $(i, j) \in I_{a,b}^2$ for some $a, b \in A$. Then $i \neq j$ if and only if there is a hole in $Q$ separating the trajectory from $f^n(p_1)$ to $f^{n+1}p_1$ and the trajectory from $f^n(p_2)$ to $f^{n+1}(p_2)$.

### 5.2. Parallel trajectories and alternating orbits

Next, we shall apply $B$-codings to parallel trajectories - a special situation where the $B$-coding can be interpreted as the coding of a single dynamical system (Corollary 5.5). The notion of ‘parallel trajectories’ is made precise in the following definition.

**Definition 5.3:** Two phase points $(x_1, \theta_1)$ and $(x_2, \theta_2)$ are said to be parallel if the following properties are satisfied.

1. The base points $x_1$ and $x_2$ are two distinct points on the same edge of $Q$ and $\theta_1 = \theta_2$.
2. The parallel phase points $(x_1, \theta_1)$ and $(x_2, \theta_2)$ both have an infinite forward orbit under $f$ and the same edge coding.

In this case, the two unfolded trajectories $\Gamma_1$ and $\Gamma_2$ in $Q^\infty$ associated with $p_1$ and $p_2$ will be called parallel trajectories from $p_1$ and $p_2$. The parallel separation between $p_1$ and $p_2$ is defined as $|x_1 - x_2| \sin \theta_1$. Geometrically, this is nothing but the perpendicular distance between $\Gamma_1$ and $\Gamma_2$.

From now on until the end of this section, let $p_1$ and $p_2$ be parallel phase points with a parallel separation $L > 0$. Assume that $p_1$ is on the left side of $p_2$. Let $\beta = (\beta_n)_{n \geq 0}$ be the $B$-coding associated with $(p_1, p_2)$. 
The alternating orbit \((P_n)_{n \geq 0}\) of the parallel phase points \(p_1\) and \(p_2\) is defined as the sequence
\[
p_1, f(p_2), f^2(p_1), f^3(p_2), f^4(p_1), \ldots .
\]
In other words, the sequence \((P_n)_{n \geq 0}\) is given by
\[
P_n = \begin{cases} 
  f^n(p_1) & \text{if } n \text{ is even} \\
  f^n(p_2) & \text{if } n \text{ is odd}
\end{cases} \quad \text{for all } n \geq 0.
\]

We would like to interpret the alternating orbit associated with \((p_1, p_2)\) as the orbit of a single point in a new dynamical system. We will also interpret the \(B\)-coding associated with parallel phase points as a natural coding of this dynamical system, which will be useful for the proof of Lemma 5.6. The main idea is to compose \(f\) with a map \(\tau\), whose job is to translate each phase point to the left by a distance calibrated by \(L\), so that the alternating orbit will be the orbit of \(p_1\) under the iterations of \(\tau \circ f\).

Since \(p_1\) and \(p_2\) are parallel phase points, by the law of reflection their images under the billiard map \(f\) are still parallel with the same parallel separation \(L\). Suppose for the moment that \(f^n(p_2)\) is the phase point on the right. If \((x_n, \theta_n)\) is the coordinates of \(f^n(p_2)\) as defined in Section 3, then the coordinates of \(f^n(p_1)\) will be given by
\[
  f^n(p_1) = \left( x_n - \frac{L}{\sin(\theta_n)}, \theta_n \right).
\]

This motivates us to define a parallel translation map \(\tau\) associated with \((p_1, p_2)\) as follows:
\[
  \tau : \quad F \to V,
  \quad (x, \theta) \mapsto \left( x - \frac{L}{\sin \theta}, \theta \right),
\]
where
\[
  F = \bigcup_{a \in A} \{ (x, \theta) \in E_a \mid (x - L/\sin \theta, \theta) \in E_a \}.
\]
is a finite union of some disjoint open subsets of \(V\). The set \(F\) is exactly the subset of \(V\) on which \(\tau(p)\) and \(p\) always lie on the same edge. Notice that the map \(\tau\) is also a homeomorphism onto its image.

This map \(\tau\) helps define another family of open subsets \(U_{i,j}^{i,j}\) in \(V\) which partitions \(\tau(F)\): within each \(V_{a,b}^i\), we define subsets \(U_{a,b}^{i,j}\) by
\[
  U_{a,b}^{i,j} = V_{a,b}^i \cap \tau(V_{a,b}^j \cap F).
\]

**Lemma 5.4:** For all \(p \in U_{a,b}^{i,j}\), we have
\[
  \tau \circ f(p) = f \circ \tau^{-1}(p).
\]

**Proof:** Let \(p \in U_{a,b}^{i,j}\) for some \(i, j \in I_{a,b}\) and \(a, b \in A\). Note that \(\tau^{-1}(p)\) and \(p\) are parallel phase points with parallel separation \(L\). Since the billiard map \(f\) preserves the parallel
Using Lemma 5.4, we can interpret the parallel trajectories from $p_1$ and $p_2$ as an orbit in a single dynamical system with a natural coding given by the $B$-coding of $p_1$ and $p_2$.

**Corollary 5.5:** The alternating orbit $(P_n)_{n \geq 0}$ is equal to the orbit of $p_1$ under $\tau \circ f$. Moreover, if $\beta$ is the $B$-coding of $(p_1, p_2)$, then for all $n \geq 0$ and $(i, j) \in \bigcup_{a, b \in A} I_{a, b} \times I_{a, b}$, we have

$$(\tau \circ f)^n(p_1) \in U_{a, b}^{ij} \text{ if and only if } \beta_n = (i, j) \in I_{a, b}.$$

**Proof:** Let $\alpha = (\alpha_n)_{n \geq 0}$ be the edge coding of $p_1$ and $p_2$. The phase point $p_1$ is initially on the left hand side of $p_2$. Since each reflection inverts the orientation, the definition of the alternating orbit ensures that $P_n$ is always on the left side of the parallel pair $f^n(p_1)$ and $f^n(p_2)$. Thus, by definition of $B$-coding, we have $P_n \in V_{\alpha_n, \alpha_{n+1}}^j$ and $\tau^{-1}(P_n) \in V_{\alpha_n, \alpha_{n+1}}^j$ where $\beta_n = (i, j) \in I_{\alpha_n, \alpha_{n+1}}$. By Lemma 5.4, we also have $P_{n+1} = \tau \circ f(P_n)$ for $n \geq 0$ and $P_0 = p_1$. This shows that $(P_n)_{n \geq 0}$ is equal to the orbit of $p_1$ under $\tau \circ f$. 

### 5.3. Construction of generalized trajectories

For the rest of this section, we apply the notion of generalized trajectories defined in Section 4 to study parallel trajectories. Let $\beta$ be the $B$-coding of parallel phase points $p_1$ and $p_2$. Let $S : B^\mathbb{N} \to B^\mathbb{N}$ be the left shift map by one index, i.e.

$$S((v_1, v_2, v_3, \ldots)) = (v_2, v_3, v_4, \ldots) \quad \text{for all } v = (v_1, v_2, v_3, \ldots) \in B^\mathbb{N}.$$
Figure 7. Illustration of Lemma 5.6.

Define $\Omega$ to be the $\omega$-limit of the action of $S$ on $\beta$, i.e.

$$\Omega = \bigcap_{n \geq 0} \{S^k(\beta) \mid k \geq n\}. \quad (5)$$

Every $\omega \in \Omega$ is the limit of a sequence $(S^i(\beta))_{i \geq 1}$ with $i_1 < i_2 < \cdots$. The following lemma shows that this approximation on the level of symbolic encodings can be converted into a geometrical result, where a unique parallel pair of generalized trajectories can be associated with $\omega$ and this pair can be approximated arbitrarily well by parts of the physical trajectories from $p_1$ and $p_2$.

Lemma 5.6: Let $\beta$ be the $B$-coding of parallel phase points $p_1$ and $p_2$. Let $\Omega$ be the $\omega$-limit of $\beta$ under $S$ as defined in (5). Then for every $\omega \in \Omega$, there exists a sequence $0 < j_1 < j_2 < \cdots$ such that the following properties hold for $q_m := (\tau \circ f)^m(p_1)$.

1. For all $m \geq 1$, the first $m$ terms of the $B$-coding associated with the parallel phase points $q_m$ and $\tau^{-1}(q_m)$ coincide with the first $m$ terms of $\omega$.
2. There exists a sequence $(z_k)_{k \geq 0}$ of generalized phase points such that for all $k \geq 0$,

$$(\tau \circ f)^k(q_m) \to z_k$$

as $m \to \infty$. This is illustrated by Figure 7.
3. The sequences

$$z_0, \tau^{-1}(z_1), z_2, \tau^{-1}(z_3), z_4, \tau^{-1}(z_5), \ldots$$

and

$$\tau^{-1}(z_0), z_1, \tau^{-1}(z_2), z_3, \tau^{-1}(z_4), z_5, \ldots$$

are two generalized trajectories.

The proof of Lemma 5.6 requires the following result, whose proof uses elementary geometry and can be found in the appendix.

Lemma 5.7: For each $a, b \in A$ and $i, j \in I_{a, b}$, the billiard map $f$ restricted onto $U^{ij}_{a, b}$ is uniformly continuous. Thus the map $f|_{U^{ij}_{a, b}}$ can be continuously extended to $\overline{U^{ij}_{a, b}}$. 
In addition, we can also extend the map $\tau : F \to V$ defined by (1) continuously to $\overline{F}$, where $F$ is the open subset of $V$ defined in (2). By Lemma 5.7, we may compose the extended map $f$ with $\tau$ to get a continuous map from $U_{a,b}^{h_{ij}}$ to $V$.

**Proof of Lemma 5.6:** It follows from the definition of $\Omega$ that there exists a sequence $0 \leq i_1 < i_2 < \cdots$ such that $(S^{n_p}(\beta))_{n \geq 1}$ converges to $\omega \in \Omega$. Hence, for all $m \geq 1$, there exists $n(m) \geq 0$ such that

$$S^{i_{n(m)}}(\beta)_k = \omega_k \quad \text{for} \ 0 \leq k \leq m.$$ 

By definition of $B$-coding and Corollary 5.5, for $0 \leq k \leq m$, the phase point $(\tau \circ f)^{i_{n(m)}+k}(p_1)$ lies in $U_{a,b}^{h_{ij}}$ where $a, b, i, j$ are such that $\omega_k = (i,j) \in I_{a,b} \times I_{a,b}$. Let us set

$$j_m := i_{n(m)}.$$ 

Then $q_m := (\tau \circ f)^{i_m}(p_1)$ satisfies (1) by construction. To simplify the notation, we also define

$$U_k = U_{a_k,b_k}^{i_k,j_k} \quad \text{where} \quad \omega_k = (i_k,j_k) \in I_{a_k,b_k} \times I_{a_k,b_k}.$$ 

We can therefore write

$$q_{m'} \in \bigcap_{k=0}^{m} (\tau \circ f)^{-k}(U_k) \quad \text{for all} \ m' \geq m.$$ 

Since $q_m$ is in the relatively compact set $U_0$ for all $m \geq 1$, by extracting a subsequence, we may assume that $q_m$ converges to a generalized phase point $z_0 \in \overline{U_0}$ as $m \to \infty$. The continuity of $\tau \circ f$, which is guaranteed by Lemma 5.7 and the remarks after it, allows us to define the generalized phase point

$$z_k := \lim_{m \to \infty} (\tau \circ f)^k(q_m) \in \overline{U_k} \quad \text{(6)}$$

for all $k \geq 1$. By construction, the sequence $(z_k)$ satisfies the required properties. In particular, we deduce from (6) and Lemma 5.4 that the sequences

$$z_0, \tau^{-1}(z_1), z_2, \tau^{-1}(z_3), z_4, \ldots \quad \text{and} \quad \tau^{-1}(z_0), z_1, \tau^{-1}(z_2), z_3, \tau^{-1}(z_4), \ldots$$

are generalized trajectories.

As an application of Lemma 5.6, we prove the following result which will be used in the proof of the main theorem in Section 6.

**Corollary 5.8:** Let $\beta$ be the $B$-coding of parallel phase points $p_1$ and $p_2$. Suppose the edge coding of $p_1$ and $p_2$ is nonperiodic. Then the $\omega$-limit set $\Omega$ of $\beta$ under the left shift map $S$ does not contain periodic elements.

In preparation for the proof of Corollary 5.8, we recall the following result on periodic edge codings in [7, Theorem 1]. This result can be readily extended to generalized trajectories.
**Figure 8.** Proof of Corollary 5.8.

**Theorem 5.9:** If a phase point \( p \) has infinite forward orbit under the billiard flow and the associated edge coding is periodic, then the orbit of \( p \) under the billiard map \( f \) is periodic.

**Proof of Corollary 5.8:** Suppose there exists a periodic \( \omega \in \Omega \). We show that the edge coding of \( p_1 \) and \( p_2 \) must be periodic. Apply Lemma 5.6 to \( \omega \) to find a subsequence \( (q_m)_{m \geq 1} \) from the sequence of phase points \( ((\tau \circ f)^n p_1)_{n \geq 1} \) and a sequence of generalized phase points \( (z_n)_{n \geq 1} \) satisfying the conclusion of Lemma 5.6. Construct the infinite corridor \( Q^\infty \) by unfolding the polygon \( Q \) according to \( \omega \).

By periodicity of \( \omega \), we have \( ST(\omega) = \omega \) for some \( T > 1 \). We may assume \( T \) is even since if \( T \) is odd it suffices to consider \( 2T \). Hence, the polygon \( QT \) in the unfolding \( Q^\infty \) is obtained from \( Q \) by an even number of reflections and thus has the same orientation as \( Q_0 \). Theorem 5.9 extended to generalized trajectories implies that the unfolding \( Q^\infty \) is periodic and \( z_n = z_{n+T} \) for all \( n \geq 0 \). In particular, we have that \( QT \) is a parallel translation of \( Q_0 \) (See Figure 8). By periodicity, the unfolding \( Q^\infty \) is obtained by joining infinitely many copies of \( QT \) end to end.

Choose any \( m \geq T \). The phase points \( f^T(q_m) \) and \( q_m \) lie on the same edge of \( Q \) by Lemma 5.6 point (1) and they are parallel. By point (2) of Lemma 5.6 and the periodicity \( z_n = z_{n+T} \), both \( f^T(q_m) \) and \( q_m \) converge to \( z_0 \) and both \( f^T(\tau^{-1}(q_m)) \) and \( \tau^{-1}(q_m) \) converge to \( \tau^{-1}(z_0) \). With a sufficiently large \( m \), we may assume that the base point of either \( f^T(q_m) \) or \( f^T(\tau^{-1}(q_m)) \) lies between the base points of \( q_m \) and \( \tau^{-1}(q_m) \).

Suppose, for example, that the base point of \( f^T(q_m) \) lies strictly between the base points of \( q_m \) and \( \tau^{-1}(q_m) \) as shown in Figure 8. Consider the four parallel physical trajectories \( S_m, l, S'_m \) and \( l' \) from \( q_m, f^T(q_m), \tau^{-1}(q_m) \) and \( f^T(\tau^{-1}(q_m)) \), respectively. If \( l \) intersects an edge \( e \) in \( Q^T_0 \), then so does \( l' \) by edge coding. This forces \( S'_m \) to intersect \( e \) since \( S'_m \) lies between \( l \) and \( l' \). Similarly, if \( S_m \) intersects an edge in \( Q^T_0 \), then \( l \) must intersect this edge, too. It follows that the first \( T \) terms of the edge coding of \( f^T(q_m) \) coincide with the first \( T \) terms of the edge coding of \( q_m \). This further implies that the base point of \( f^{2T}(q_m) \) lies between the base points of \( f^T(q_m) \) and \( f^T(\tau^{-1}(q_m)) \).

By induction on \( n \geq 1 \), we can deduce that the first \( T \) terms of the edge codings of \( f^{nT}(q_m) \) and \( f^{(n+1)T}(q_m) \) coincide, and in particular, the edge coding of \( q_m \) is periodic with period \( \leq T \). By Theorem 5.9, the physical trajectory from \( q_m \) is periodic, and thus the edge coding associated to \( p_1 \) and \( p_2 \) is also periodic.

**6. Proof of the main theorem**

In this section, we take on the proof of the main theorem.
**Main Theorem:** Let \( Q \) be a polygon such that all the holes of \( Q \) have nonzero minimal diameters. Suppose the edges of \( Q \) are indexed by \( \mathcal{A} \). Then for any nonperiodic sequence \( \alpha \) in \( \mathcal{A} \), the set \( X(\alpha) \) contains at most one point, i.e. there is at most one point \( p \in V \) whose edge coding is equal to \( \alpha \).

The starting point of the proof is the observation that \( X(\alpha) \) consists of parallel phase points. This well-known result is proven, for example, in [7, Lemma 1].

**Lemma 6.1:** If two phase points \((x, \theta)\) and \((x', \theta')\) with infinite forward orbit under \( f \) do not have the same direction, that is \( \theta \neq \theta' \), then they do not have the same edge coding.

Let us assume \( Q \) is a polygon satisfying the condition of the main theorem. Let \( \delta > 0 \) be smaller than the minimal diameter of every hole in \( Q \). Our proof of the main theorem proceeds by contradiction. Suppose towards a contradiction that there exists at least one nonperiodic \( \alpha \in \mathcal{A}^\mathbb{N} \) with at least two distinct phase points in \( X(\alpha) \). In view of Lemma 6.1, all phase points in \( X(\alpha) \) are necessarily parallel.

Under the above assumption, we now choose \( \epsilon > 0 \) and parallel phase points \( p \) and \( q \) with a parallel separation \( L > 0 \) satisfying the following properties.

1. The two edge codings of \( p \) and \( q \) are identical and nonperiodic.
2. The edge coding of any two parallel phase points with parallel separation \( \geq L + \frac{\epsilon}{4} \) is periodic.
3. \( \epsilon \) is smaller than both \( L/2 \) and \( \delta/2 \).

This choice is possible as the parallel separation is bounded by the maximal length of the edges of \( Q \). Assume, without loss of generality, that \( p \) is on the left side of \( q \). Let \( \tau \) be the left translation map associated to \((p, q)\) as defined in Section 5. Then \( q = \tau^{-1}(p) \). Let \( \beta \in \mathcal{B}^\mathbb{N} \) be the \( \mathcal{B} \)-coding of \( p \) and \( q \). As defined in (5), we let \( \Omega \) denote the \( \omega \)-limit set of \( \beta \) under the left shift map \( S \). Recall the following classical theorem due to Birkhoff (see, for example, [6, Theorem 1.16])

**Theorem 6.2:** If \( Z \) is compact and \( T : Z \rightarrow Z \) is continuous, then the topological dynamical system \((Z, T)\) contains a uniformly recurrent point \( z \in Z \), i.e. for all open neighbourhood \( V \) of \( z \), the set

\[
\{ n \in \mathbb{N} \mid T^n(z) \in V \}
\]

can be arranged as an increasing sequence \( s_1 < s_2 < s_3 < \cdots \) with bounded gaps \( s_{n+1} - s_n \).

Applying Theorem 6.2 with \( Z = \Omega \) and \( T = S \), we obtain a uniformly recurrent point \( \omega \in \Omega \) under \( S \). Since \( p \) and \( q \) are assumed to have nonperiodic edge coding, by Corollary 5.8, the point \( \omega \) is a nonperiodic sequence. By Lemma 5.6 applied on \( \omega \in \Omega \), there exist a sequence \( (q_m = (\tau \circ f)^m(p))_{m \geq 1} \) of phase points and a sequence \( (z_n)_n \) of generalized phase points such that the sequences

\[
z_0, \tau^{-1}(z_1), \tau^{-1}(z_2), \tau^{-1}(z_3), \tau^{-1}(z_4), \tau^{-1}(z_5), \ldots \quad \text{(L)}
\]

and

\[
\tau^{-1}(z_0), z_1, \tau^{-1}(z_2), z_3, \tau^{-1}(z_4), z_5, \ldots \quad \text{(R)}
\]
are two generalized trajectories which can be approximated arbitrarily well in the sense of Lemma 5.6 point (2) by parallel trajectories from \( q_m \) and \( \tau^{-1}(q_m) \) as \( m \to \infty \). Let \( \beta(m) \) denote the \( \mathcal{B} \)-coding of \( q_m \) and \( \tau^{-1}(q_m) \). Note that \( \beta(m) = S^m(\beta) \).

### 6.1. Locating vertices in \( Q^\infty \)

When \( Q \) is unfolded according to \( \omega \), the two sequences \((L)\) and \((R)\) define two generalized trajectories \( \Gamma_L \) and \( \Gamma_R \), respectively, in \( Q^\infty \). These two trajectories satisfy properties in Remark 4.2. Note that the parallel separation between \( \Gamma_L \) and \( \Gamma_R \) is the same as the parallel separation \( L \) between \( p \) and \( q \).

Let us also define \( S_m \) and \( S'_m \) to be the parallel billiard trajectories from \( q_m \) and \( \tau^{-1}(q_m) \). For \( m \geq 1 \), let \( N \geq 1 \) be the largest integer depending on \( m \) such that \( \beta(m)_n = \omega_n \) for \( n = 0, 1, 2, \ldots, N - 1 \). The index \( N \) is the smallest index at which \( \beta(m) \) and \( \omega \) differ, and the four trajectories \( S_m, S'_m, \Gamma_L \) and \( \Gamma_R \) intersect the same sequence of edges in the finite corridor \( Q^N_0 \). It follows from point (1) of Lemma 5.6 that \( N \geq m \). We have the following lemma describing the location of some vertices in \( Q^N_0 \).

**Lemma 6.3**: Suppose that \( \Gamma_L \) and \( \Gamma_R \) enter \( Q_N \) via the edge \( e \) and exit \( Q_N \) via the edge \( e' \). In other words, the edge \( e \) is the boundary between \( Q_{N-1} \) and \( Q_N \) and \( e' \) is the boundary between \( Q_N \) and \( Q_{N+1} \). Let \( \beta(m)_N = (i, j) \in I_{a,b} \times I_{a,b} \) and \( \omega_N = (i', j') \in I_{a,b'} \times I_{a,b'} \) where \( a, b, b' \in \mathcal{A} \). Note that \( e = e_a \) and \( e' = e_{b'} \). Within the finite corridor \( Q^N_0 \), let \( T_1 \) be the interior of the region bounded by the straight lines \( S_m, \Gamma_L, e \) and \( T_2 \) be the interior of the region bounded by the straight lines \( S'_m, \Gamma_R, e \). The regions \( T_1 \) and \( T_2 \) are illustrated in Figure 9. The following statements hold.

1. The regions \( T_1 \) and \( T_2 \) do not contain any vertices.
2. Suppose \( S_m \) bounds \( T_1 \) from the left. No vertex in \( \overline{T_1} \cap \Gamma_L \) blocks \( \Gamma_L \) from the left and no vertex in \( \overline{T_2} \cap \Gamma_R \) blocks \( \Gamma_R \) from the left. Analogous statement holds if \( S_m \) bounds \( T_1 \) from the right.
3. If \( b = b' \), then there exists a vertex \( v \) in \( Q_N \) lying strictly between \( S_m \) and \( \Gamma_L \) or between \( S'_m \) and \( \Gamma_R \). Moreover, the vertex \( v \) is attached to a hole of \( Q \).
4. If \( b \neq b' \), then the base points of \( f^{N+1}(q_m) \) and \( f^{N+1}(\tau^{-1}(q_m)) \) either both lie on the right side of \( \Gamma_R \) or both lie on the left side of \( \Gamma_L \).

**Proof**: It follows from the choice of \( N \) and the definition of \( \mathcal{B} \)-codings that the first \( N \) terms of the \( \mathcal{V} \)-codings associated to \( S_m \) and \( \Gamma_L \) coincide. By Lemma 3.3, there are no holes between the trajectories \( S_m \) and \( \Gamma_L \) before they meet the edge \( e \). Also, there must not be any vertices of the outer boundary of \( Q_n \) with \( n = 0, 1, \ldots, N - 1 \) which lies strictly between \( S_m \) and \( \Gamma_L \), for, if not, the interior of a nonreflecting edge would transversally intersect either \( S_m \) or \( \Gamma_L \), contradicting Remark 4.2. A similar argument applies to \( S'_m \) and \( \Gamma_R \). This implies (1).

The statement (2) follows immediately from the first statement of Lemma 4.4 and the assumption that the \( \mathcal{B} \)-coding of \( (S_m, S'_m) \) coincide with the \( \mathcal{B} \)-coding of \( (\Gamma_L, \Gamma_R) \) up to the \( N \)th term.
If \( b = b' \), then the four trajectories \( S_m, S'_m, \Gamma_L \), and \( \Gamma_R \) exit \( Q_N \) by the same edge \( e' \). There is a hole in \( Q_N \) lying either between \( S_m \) and \( \Gamma_L \) or between \( S'_m \) and \( \Gamma_R \). If \( b \neq b' \), then \( S_m \) and \( S'_m \) exit \( Q_N \) by different edges from \( e' \).

The analogous statements for \( v \) lying on \( \Gamma_R \) also hold.
**Proof:** Suppose \( v \) lies on \( \Gamma_L \) and \( v \) is an end vertex of some reflecting edge \( e \). Since \( \Gamma_R \) intersects \( e \) on the right side of \( \Gamma_L \) and \( \Gamma_R \) has a nonzero parallel separation from \( \Gamma_L \), the edge \( e \) cannot overlap \( \Gamma_L \) or lie on the left side of \( \Gamma_L \). Consequently, the interior of \( e \) lies on the right side of \( \Gamma_L \) and thus \( v \) blocks \( \Gamma_L \) from the left by Definition 4.3.

Suppose \( v \) is not an end vertex of any reflecting edge. If \( v \) does not block \( \Gamma_L \) either from left or from right, then all edges containing \( v \) must overlap \( \Gamma_L \). Since \( Q \) has nonzero diameter and all the holes of \( Q \) have nonzero diameters, the connected component of \( \partial Q \) containing \( v \) is not contained in \( \Gamma_L \). Thus, there exists a vertex \( u \) connected to \( v \) via a sequence of edges lying in \( \Gamma_L \) such that \( u \) is the end vertex of an edge \( e' \) not overlapping \( \Gamma_L \). Since \( e' \) lies on either the left side or the right side of \( \Gamma_L \), we see that \( u \) blocks \( \Gamma_L \) from left or from right. 

### 6.2. Existence of uniformly recurrent vertices

Recall that Corollary 5.8 and the nonperiodicity of the edge coding of \( p \) and \( q \) imply that the sequence \( \omega \) is nonperiodic. The main objective of this subsection is to establish a useful consequence of Theorem 6.2: the existence of certain uniformly recurrent vertices near the generalized trajectories (Corollary 6.8).

For convenience, we define the following subsets of \( Q^\infty \). Let \( \eta > 0 \).

- The set \( \mathcal{L}_L(\eta) \) is the set of vertices in \( Q^\infty \) which either lies in the open left \( \eta \)-strip of \( \Gamma_L \) or blocks \( \Gamma_L \) from the left.
- The set \( \mathcal{R}_R(\eta) \) is the set of vertices in \( Q^\infty \) which either lies in the open right \( \eta \)-strip of \( \Gamma_R \) or blocks \( \Gamma_R \) from the right.
- The set \( \mathcal{R}_L(\eta) \) is the set of vertices in \( Q^\infty \) which either lies in the open right \( \eta \)-strip of \( \Gamma_L \) or blocks \( \Gamma_L \) from the right.
- The set \( \mathcal{L}_R(\eta) \) is the set of vertices in \( Q^\infty \) which either lies in the open left \( \eta \)-strip of \( \Gamma_R \) or blocks \( \Gamma_R \) from the left.

These definitions are motivated by our results in Lemma 4.4: for example, if \( \mathcal{L}_L(\eta) \) is empty, then by Lemma 4.4 there exists a physical trajectory in the open left \( \eta \)-strip having the same edge coding as \( \Gamma_L \).

In Lemma 6.5 and Proposition 6.6, we exploit the fact that \( L \) is \( \epsilon/4 \) away from the supremum of parallel separations between nonperiodic parallel trajectories. The basic idea is as follows. The choice of \( \epsilon \) ensures that no pairs of trajectories with parallel separations \( \geq L + \epsilon/4 \) have the same edge coding as \( \Gamma_L \). Thus, there must be vertices in the sets \( \mathcal{L}_L(\eta) \), \( \mathcal{R}_R(\eta) \), \( \mathcal{L}_R(\eta) \) and \( \mathcal{R}_L(\eta) \) for some small \( \eta > 0 \) to prevent the existence of such parallel trajectories.

Let us put \( \epsilon' = \epsilon/2 \) and let \( \mathcal{L}_L = \mathcal{L}_L(\epsilon') \), \( \mathcal{R}_R = \mathcal{R}_R(\epsilon') \), \( \mathcal{L}_R = \mathcal{L}_R(\epsilon') \) and \( \mathcal{R}_L = \mathcal{R}_L(\epsilon') \).

**Lemma 6.5:** \( \mathcal{L}_L \cup \mathcal{R}_R \) contains infinitely many vertices.

**Proof:** Suppose not, then \( \mathcal{L}_L \cup \mathcal{R}_R \) is contained in \( Q^M_0 \) for a sufficiently large \( M \) as illustrated in Figure 10. By only considering the parts of \( \Gamma_L \) and \( \Gamma_R \) after the \( M \)th reflection, we
If \( L \cup R \) contains only finitely many vertices, it will be contained in \( Q_0^M \) for sufficiently large \( M \). In \( Q_\infty \), there are physical trajectories \( l_1 \) and \( l_2 \) parallel to \( \Gamma_L \) and \( \Gamma_R \) with parallel separation arbitrarily close to \( L + 2\epsilon' \).

Figure 10. Illustration of three cases in Proposition 6.6.

Proposition 6.6: At least one of the following statements holds:

1. Both \( L \) and \( R \) contain infinitely many vertices.
2. Both \( L \) and \( L \) contain infinitely many vertices.
3. Both \( R \) and \( R \) contain infinitely many vertices.

The sets involved in these three cases are illustrated in Figure 11.

Proof: Suppose (1) does not hold. By Lemma 6.5, either \( L \) or \( R \) is finite but not both. We first assume \( R \) is finite and show that \( L \) is infinite, thereby deducing (2). Assume for a contradiction that \( L \) is finite. Then by starting \( \Gamma_L \) and \( \Gamma_R \) at a later time, we can assume that \( R \cup L \) is empty. By Lemma 4.4, there exists a physical trajectory \( l \) in the open right \( \epsilon' \)-strip of \( \Gamma_R \) and a physical trajectory \( l' \) in the open right \( \epsilon' \)-strip of \( \Gamma_L \) such that \( l \) and \( l' \) have the same edge coding as \( \Gamma_L \) and \( \Gamma_R \). Moreover, the parallel separation between \( l \) and \( l' \) can be chosen to be arbitrarily close to \( L + \frac{\epsilon}{2} \). This contradicts the choice of \( \epsilon \) in the beginning of the proof and the nonperiodicity of \( \omega \). Therefore, \( R \) is infinite and we have (2). A similar argument shows that if \( L \) is finite, then we will have (3).

For convenience, we make the following definition.
Both (3) In particular, both
If (1) in Proposition 6.6 holds, then we can pick
Proof:
S
QM
Both
L
R
Lemma 4.4 implies that
We can show
different sides of the parallel trajectories.3 Due to the length of
edges connecting
either from left or from right, and
As u b s e t
Definition 6.7: A subset E of \( Q^\infty \) is said to be \textit{uniformly recurrent} if there exists an
increasing sequence \( i_1 < i_2 < i_3 < \cdots \) with uniformly bounded gaps \( i_{n+1} - i_n \) such that
\( Q^{i_n} \cap E \neq \emptyset \) for all \( n \geq 1 \).

Corollary 6.8: At least one of the following statements holds:

1. Both \( L_L(\epsilon) \) and \( R_R(\epsilon) \) are uniformly recurrent.
2. Both \( L_L(\epsilon) \) and \( R_L(\epsilon) \) are uniformly recurrent.
3. Both \( R_R(\epsilon) \) and \( R_R(\epsilon) \) are uniformly recurrent.

In particular, both \( L_L(\epsilon) \cup L_R(\epsilon) \) and \( R_L(\epsilon) \cup R_R(\epsilon) \) are uniformly recurrent.

Proof: If (1) in Proposition 6.6 holds, then we can pick \( v_1, v_2 \) and \( v_3 \) in \( Q^\infty \), appearing
in that order, such that \( v_1, v_3 \in L_L(\epsilon/2) \) and \( v_2 \in R_R(\epsilon/2) \). Let \( u \) be the closest point on
the segment \( v_1v_3 \) to \( v_2 \). By choosing \( v_1 \) and \( v_3 \) sufficiently far apart, we can assume the
line segment \( uv_2 \) is approximately perpendicular to \( \Gamma_L \) and \( \Gamma_R \) so that the length of \( uv_2 \) is
less than \( \epsilon + L \). Let \( Q_0^M \) be the finite corridor containing these three vertices as shown in
Figure 12. By the uniform recurrence property of \( \omega \), the trajectories \( \Gamma_L \) and \( \Gamma_R \) will pass
through copies of \( Q_0^M \) infinitely many times with bounded time interval. We will deduce
case (1) by showing that each copy of \( Q_0^M \) contains a vertex in \( R_R(\epsilon) \) and a vertex in \( L_L(\epsilon) \).
Let us denote one of these copies of \( Q_0^M \) by \( \tilde{Q}_0^M \) and denote the images of \( v_i \) in \( \tilde{Q}_0^M \) by \( \tilde{v}_i \) for
\( i = 1, 2, 3 \).

By the \( B \)-coding of \( \Gamma_L \) and \( \Gamma_R \), the line segment \( \tilde{v}_1\tilde{v}_3 \) and the vertex \( \tilde{v}_2 \) will lie on two
different sides of the parallel trajectories.3 Due to the length of \( uv_2 \), the vertex \( \tilde{v}_2 \) will lie
at most \( \epsilon \) away from one of the two trajectories and \( \tilde{v}_1 \) or \( \tilde{v}_3 \) will lie at most \( \epsilon \) away from
the other trajectory. If \( \tilde{v}_1 \) does not lie on \( \Gamma_L \) or \( \Gamma_R \), then we have \( \tilde{v}_1 \in R_R(\epsilon) \) or \( \tilde{v}_1 \in L_L(\epsilon) \).
If \( \tilde{v}_1 \) lies on \( \Gamma_L \), then by Lemma 6.4 there exists a vertex \( \tilde{w} \) in \( \Gamma_L \cap \tilde{Q}_0^M \) which blocks \( \Gamma_L \)
either from left or from right, and \( \tilde{w} \) is either equal to \( \tilde{v}_1 \) or connected to \( \tilde{v}_1 \) via a sequence
of edges \( e_1, e_2, \ldots \) overlapping \( \Gamma_L \). Clearly, these edges \( e_1, e_2, \ldots \) are nonreflecting edges.
We can show \( \tilde{w} \in L_L(\epsilon) \) by proving that \( \tilde{w} \) does not block \( \Gamma_L \) from the right as follows.

We first consider the case \( v_1 \in L_L(\epsilon) \). Choose a physical trajectory \( S \) in \( Q_0^M \) with the
same \( V \)-coding as \( \Gamma_L \cap Q_0^M \). It follows from Lemma 4.4 that \( v_i \) must lie on the left side
of \( S \). Let \( w \in Q_0^M \) be the vertex whose image in \( \tilde{Q}_0^M \) is \( \tilde{w} \). Since \( S \) does not intersect the
edges connecting \( w \) and \( v_i \), the vertex \( w \) is also on the left side of \( S \). Consider the image
\( \tilde{S} \) of \( S \) in \( \tilde{Q}_0^M \). Since \( \tilde{w} \) lies on the left side of \( \tilde{S} \) and \( \tilde{S} \) has the same \( V \)-coding as \( \Gamma_L \cap Q_0^M \),
Lemma 4.4 implies that \( \tilde{w} \) does not block \( \Gamma_L \) from the right. The argument for case \( v_1 \in R_R(\epsilon) \) is similar except that \( \tilde{S} \) is taken to have the same \( V \)-coding as \( \Gamma_R \cap Q_0^M \). In this case,
both \( v_1 \) and \( w \) lie on the right side of \( S \), and \( \tilde{w} \) lies on the left side of \( \tilde{S} \). The same conclusion follows.

A similar argument applies if \( \tilde{v}_1 \) lies on \( \Gamma_R \) and we can show that a vertex \( \tilde{w} \in \tilde{Q}_0^M \) lies in \( \mathcal{R}_R(\epsilon) \). In conclusion, each copy \( \tilde{Q}_0^M \) contains at least one vertex in \( \mathcal{R}_R(\epsilon) \) and one vertex in \( \mathcal{L}_L(\epsilon) \). By the uniform recurrence property, we have case (1).

If (2) in Proposition 6.6 holds, then we can choose \( M \geq 0 \) and vertices \( v_1, v_2 \) and \( v_3 \) in \( Q_0^M \), appearing successively in that order, such that \( v_1, v_3 \in \mathcal{L}_L(\epsilon/2) \) and \( v_2 \in \mathcal{R}_L(\epsilon/2) \) and the line segment \( uv_2 \) defined above has length less than \( \epsilon \). Again, let \( \tilde{Q}_0^M \) be one of the uniformly recurrent copies of \( Q_0^M \) in the unfolding \( Q^\infty \). Then, due to the length of \( uv_2 \), the image of \( v_2 \) in \( \tilde{Q}_0^M \) will lie at most \( \epsilon \)-away from one side of \( \Gamma_L \) and the image of \( v_1 \) or \( v_3 \) in \( Q_0^M \) will lie at most \( \epsilon \)-away on the other side of \( \Gamma_L \). We can deduce using an argument similar to case (1) above that each copy \( \tilde{Q}_0^M \) intersects both \( \mathcal{L}_L(\epsilon) \) and \( \mathcal{R}_L(\epsilon) \) nontrivially, and thus both \( \mathcal{L}_L(\epsilon) \) and \( \mathcal{R}_L(\epsilon) \) are uniformly recurrent.

Case (3) is symmetric to (2) and we omit the proof.

### 6.3. Consequences of Lemma 6.3 and Corollary 6.8

We are now able to finish the proof of the main theorem in Section 6 using Lemma 6.3 and Corollary 6.8.

By Corollary 6.8, both \( \mathcal{L}_L \cup \mathcal{L}_R \) and \( \mathcal{R}_L \cup \mathcal{R}_R \) are uniformly recurrent. Let the time intervals of recurrence in \( \mathcal{L}_L \cup \mathcal{L}_R \) and in \( \mathcal{R}_L \cup \mathcal{R}_R \) be both bounded from above by \( D > 0 \), and let \( \text{diam}(Q) \) be the diameter of \( Q \). Let \( z_0 = (x', \theta') \). Take \( m > 0 \) sufficiently large such that, with \( q_m = (x_m, \theta_m) \),

\[
(D + 2 \text{diam}(Q)) \tan(\theta_m - \theta') + \epsilon < \min \left( \frac{L}{\sin(\theta_m - \theta')}, \delta \right) \tag{♠}
\]

The choice of \( m \) is possible as \( \epsilon > 0 \) has been chosen to satisfy \( \epsilon < \min(L/2, \delta/2) \). Informally, the inequality (♠) ensures that \( S_m \) and \( \Gamma_L \) are approximately parallel such that the distance between \( S_m \) and \( \Gamma_L \) does not increase too fast with \( n \).

Recall that \( \beta(m) \) is defined as the \( B \)-coding associated with \( q_m \) and \( \tau^{-1}(q_m) \). Let \( N \geq 1 \) be the largest integer such that \( \beta(m)_i = o_i \) for \( n = 0, 1, 2, \ldots, N - 1 \). We re-use notations introduced in Lemma 6.3 and consider the two cases in (3) and (4) of Lemma 6.3 separately. We will show that both cases lead to a contradiction.

In the case where \( b = b' \), there is a hole in \( Q_N \) lying between \( S_m \) and \( \Gamma_L \) or lying between \( S'_m \) and \( \Gamma_R \). As illustrated by Figure 13(a), since the minimal diameter of the hole is strictly greater than \( \delta \), there exists some point in \( S_m \cap Q_N \) whose distance to \( \Gamma_L \) is greater than \( \delta \), which is in turn greater than \( (D + 2 \text{diam}(Q)) \tan(\theta_m - \theta') + \epsilon \) by (♠). Thus, the perpendicular distance from some point in \( S_m \cap Q_N \) to \( \Gamma_L \) must be greater than \( D \tan(\theta_m - \theta') + \epsilon \).

Recall that both \( \mathcal{L}_L(\epsilon) \cup \mathcal{L}_R(\epsilon) \) and \( \mathcal{R}_L(\epsilon) \cup \mathcal{R}_R(\epsilon) \) are uniformly recurrent by Corollary 6.8. Suppose \( Q_N \cap S_m \) lies on the left side of \( \Gamma_L \). Observe that there is an open left \( \epsilon \)-strip of \( \Gamma_L \) of length \( \geq D \) lying in \( T_1 \) and an open left \( \epsilon \)-strip of \( \Gamma_R \) of length \( \geq D \) lying in \( T_2 \) indicated by the shaded regions in Figure 13(a). Since \( \mathcal{L}_L(\epsilon) \cup \mathcal{L}_R(\epsilon) \) is uniformly recurrent, there will be a vertex lying in the closure of the two strips. In particular,
this vertex belongs to one of the following four sets:

$$\mathcal{T}_1, \quad \mathcal{T}_2, \quad \overline{\mathcal{T}_1} \cap \Gamma_L, \quad \overline{\mathcal{T}_2} \cap \Gamma_R.$$ 

However, by (1) of Lemma 6.3, this vertex cannot lie in $\mathcal{T}_1$ or $\mathcal{T}_2$. On the other hand, if the vertex lies in $\overline{\mathcal{T}_1} \cap \Gamma_L$ or $\overline{\mathcal{T}_2} \cap \Gamma_R$, then by the definitions of $\mathcal{L}_L(\epsilon)$ and $\mathcal{L}_R(\epsilon)$ this vertex blocks $\Gamma_L$ or $\Gamma_R$ from the left. This contradicts (2) of Lemma 6.3. If $Q_N \cap S_m$ lies on the right side of $\Gamma_L$ instead, then we can deduce a contradiction by applying a similar argument using the uniform recurrence of $\mathcal{R}_L(\epsilon) \cup \mathcal{R}_R(\epsilon)$ instead of $\mathcal{L}_L(\epsilon) \cup \mathcal{L}_R(\epsilon)$. Hence, the case $b = b'$ leads to a contradiction.

Next, suppose $b \neq b'$. Then $S_m$ and $S'_m$ hit a different edge from $\Gamma_R$ and $\Gamma_L$ as illustrated by Figure 13(b). Since the base points of $f^N(q_m)$ and $f^N(\tau^{-1}(q_m))$ either both lie on the left side of $\Gamma_L$ or both lie on the right side of $\Gamma_R$, the perpendicular distance from some point on $S_m$ to $\Gamma_L \cap Q_N$ must be greater than $\frac{L}{\sin(\theta_m - \theta')}$, which is in turn greater than $(D + 2 \text{diam}(Q)) \tan(\theta_m - \theta') + \epsilon$ by (♣). Now, we can use exactly the argument in the previous paragraph to deduce a contradiction in the case $b \neq b'$. The proof by contradiction is complete.

**Notes**

1. Recall that parallel phase points have the same edge coding by definition.
2. See Definition 4.3.
3. Note that the vertex appearing on the left of $\Gamma_L$ need not always be $\tilde{v}_1$ or $\tilde{v}_3$. It could also be $\tilde{v}_2$ since the orientation of $Q$ is inverted after each reflection.

**Acknowledgments**

I would like to thank Prof. Serge Troubetzkoy for his guidance and his insightful suggestions during many discussions, without which this work would not have been possible. I would also like to thank *Institut de Mathématiques de Marseille* for its conducive environment and its support during my stay.
Special thanks should also go to the referee whose suggestions had substantially improved the clarity of this article.

**Disclosure statement**

No potential conflict of interest was reported by the author(s).

**References**

[1] Bobok J. and Troubetzkoy S., *Does a billiard orbit determine its (polygonal) table?* Fund. Math. 212 (2011), pp. 129–144.

[2] Bobok J. and Troubetzkoy S., *Code and order in polygonal billiards*, Topol. Appl. 159(1) (2012), pp. 236–247.

[3] Bobok J. and Troubetzkoy S., *Homotopical rigidity of polygonal billiards*, Topol. Appl. 173 (2014), pp. 308–324.

[4] Calderon A., Coles S., Davis D., Lanier J., and Oliveira A., *How to hear the shape of a billiard table*, preprint (2018). Available at arXiv:1806.09644.

[5] Duchin M., Erlandsson V., Leininger C.J., and Sadanand C., *The growth rate for the number of singular and periodic orbits for a polygonal billiard*, Commun. Math. Phys. 169(3) (1995 May), pp. 463–473.

[6] Galperin G., Krüger T., and Troubetzkoy S., *Local instability of orbits in polygonal and polyhedral billiards*, Commun. Math. Phys. 111(1) (1987), pp. 151–160.

[7] Galperin G., Krüger T., and Troubetzkoy S., *The growth rate for the number of singular and periodic orbits for a polygonal billiard*, Commun. Math. Phys. 111(1) (1987), pp. 151–160.

[8] Katok A.B., *The growth rate for the number of singular and periodic orbits for a polygonal billiard*, Commun. Math. Phys. 111(1) (1987), pp. 151–160.

[9] Zemljakov A.N. and Katok A.B., *Topological transitivity of billiards in polygons*, Mat. Zametki 18(2) (1975), pp. 291–300.

**Appendix. Proof of Lemma 5.7**

The family of disjoint open subsets \( \{E_a\}_{a \in \mathcal{A}} \) has been defined in Section 3. We fix a metric \( d \) on \( E_a \) by defining \( d(\rho, \rho') = |x - x'| + |\theta - \theta'| \) for \( \rho = (x, \theta), \rho' = (x', \theta') \in E_a \). We need to show the uniform continuity of the billiard map \( f \) on each \( U^{ij}_{a,b} \). In fact, we will prove that \( f|_{U^{ij}_{a,b}} \) is \( M \)-Lipschitz for some \( M > 0 \) depending only on the geometry of \( Q \).

For \( a, b \in \mathcal{A} \) and \( i, j \in I_{a,b} \), fix an arbitrary \( (x, \theta) \in U^{ij}_{a,b} \) and suppose \( f(x, \theta) = (y, \phi) \in E_b \). We first calculate the coordinates of \( f(x + \epsilon_1, \theta + \epsilon_2) \).

Consider a phase point \( (x - \epsilon_1, \theta) \) in \( U^{ij}_{a,b} \) parallel to \( (x, \theta) \). Note that \( f(x, \theta) \) and \( f(x - \epsilon_1, \theta) \) land on the same edge \( e_b \) and remain parallel as shown in Figure A1(a). Let \( f(x - \epsilon_1, \theta) = (y + \delta_1, \phi) \) for some \( \delta_1 \) depending on \( \epsilon_1 \). By sine rule, we deduce that \( \delta_1 = (\sin \theta / \sin \phi) \epsilon_1 \). Therefore,

\[
f(x + \epsilon_1, \theta) = \left( y - \frac{\sin \theta}{\sin \phi} \epsilon_1, \phi \right) .
\] (A1)

Next, consider \( (x, \theta + \epsilon_2) \in U^{ij}_{a,b} \), a phase point having the same base point as \( (x, \theta) \) but a possibly different direction as illustrated in Figure A1(b). Let \( d \) be the spatial distance on \( Q \) between the base point \( x \) and the base point \( y \), we have \( f(x, \theta + \epsilon_2) = (y + \delta_2, \phi - \epsilon_2) \) for some \( \delta_2 \) depending on \( \epsilon_2 \). By sine rule, we have \( \delta_2 = d \sin(\epsilon_2) / \sin(\phi - \epsilon_2) \). Therefore,

\[
f(x, \theta + \epsilon_2) = \left( y + \frac{d \sin(\epsilon_2)}{\sin(\phi - \epsilon_2)}, \phi - \epsilon_2 \right) .
\] (A2)
Figure A1. Proof of Lemma 5.7. The edge below is $e_a$ and the edge on top is $e_b$.

Now consider a phase point $(x + \epsilon_1, \theta + \epsilon_2) \in U_{a,b}^{ij}$ whose distance from $(x, \theta)$ is $\epsilon = \epsilon_1 + \epsilon_2$. By combining (A1) with (A2), we deduce

$$f(x + \epsilon_1, \theta + \epsilon_2) = \left( y - \frac{\epsilon_1 \sin(\theta + \epsilon_2)}{\sin(\phi - \epsilon_2)} + \frac{d \sin \epsilon_2}{\sin(\phi - \epsilon_2)}, \phi - \epsilon_2 \right).$$

Therefore, the distance between $f(x, \theta)$ and $f(x + \epsilon_1, \theta + \epsilon_2)$ is explicitly given by

$$d(f(x, \theta), f(x + \epsilon_1, \theta + \epsilon_2)) = \left| \frac{\epsilon_1 \sin(\theta + \epsilon_2)}{\sin(\phi - \epsilon_2)} - \frac{d \sin \epsilon_2}{\sin(\phi - \epsilon_2)} \right| + |\epsilon_2|. \tag{A3}$$

The value of $d > 0$ is uniformly bounded from above by the diameter of the polygon $Q$. On the other hand, since both $f(x + \epsilon_1, \theta + \epsilon_2)$ and $\tau \circ f(x + \epsilon_1, \theta + \epsilon_2)$ are on the same edge $e_b$ by the definition of $U_{a,b}^{ij}$, the distance $|L/\sin(\phi - \epsilon_2)|$ between their base points is bounded from above by the length of $e_b$. In particular, the value of $|\sin(\phi - \epsilon_2)|^{-1}$ is bounded from above by the length of $e_b$ divided by $L$, and we have

$$\left| \frac{\epsilon_1 \sin(\theta + \epsilon_2)}{\sin(\phi - \epsilon_2)} \right| + \left| \frac{d \sin \epsilon_2}{\sin(\phi - \epsilon_2)} \right| \leq M'(|\epsilon_1| + |\epsilon_2|)$$

for some $M' > 0$ depending only on $L$, the length of $e_b$ and the diameter of $Q$. Therefore, the distance (A3) is not greater than $M'(|\epsilon_1| + |\epsilon_2|) + |\epsilon_2|$. This proves that $f$ is $M$-Lipschitz with $M = M' + 1$. 