N. CHERNOV
Department of Mathematics
University of Alabama at Birmingham
Birmingham, AL 35294
E-mail: chernov@math.uab.edu

G. GALPERIN
Department of Mathematics
Eastern Illinois University
Charleston, IL 61920
E-mail: cfgg@ux1.cts.eiu.edu

SEARCH LIGHT IN BILLIARD TABLES

Received April 7, 2003

DOI: 10.1070/RD2003v008n02ABEH000239

We investigate whether a search light, S, illuminating a tiny angle ("cone") with vertex A inside a bounded region $Q \in \mathbb{R}^2$ with the mirror boundary ∂Q, will eventually illuminate the entire region Q. It is assumed that light rays hitting the corners of Q terminate. We prove that: **(I)** if $Q =$ a *circle* or an *ellipse*, then either the entire Q or an annulus between two concentric circles/confocal ellipses (one of which is ∂Q) or the region between two confocal hyperbolas will be illuminated; **(II)** if $Q =$ a *square*, or **(III)** if $Q =$ a *dispersing* (Sinai) or semidespirsing billiards, then the entire region Q is will be illuminated.

1. Introduction

Let Q be a bounded region on the plane with smooth or piecewise smooth boundary ∂Q. We consider Q to be a billiard table, inside which the billiard point-like ball moves with the unit speed along a straight line segment and reflects from the boundary ∂Q according to the billiard law "the angle of reflection equals the angle of incidence". If the boundary ∂Q is not smooth, i.e., has "corners", then the billiard trajectories hitting corners terminate. We can also think of the billiard trajectory as of a light ray reflecting off the mirror boundary ∂Q; we will call Q a (2-dimensional) *room* in this case. (Thus, the notions *billiard table* and *room* will be treated as synonyms throughout this article.) All the points in the path of the light are considered illuminated.

Clearly, one light ray, or finitely many rays, or even a countable set of light rays cannot illuminate the whole table Q entirely. Indeed, any smooth curve inside Q can be intersected by the light rays only in a countable set of points, whereas the curve contains uncountable set of points; thus, uncountably many points in each area inside Q are not illuminable. However, in many cases, one light ray can illuminate the region Q everywhere densely, e.g., if Q is a square, equilateral triangle, etc.

Consider now a point light source S inside the room Q that emits rays in all directions which are reflected by the mirror walls of the room. *Will all the points in the room be illuminated after many reflections?* In other words, is any region Q illuminable? This problem is known in the literature as the "Illumination Problem". It has been tentatively traced back to Ernst Straus in early 1950's. Two questions were posed:

(1) Is a region illuminable from **every** point in the region? and

(2) Is a region illuminable from at least **one** point in the region?

In 1958, L. Penrose and R. Penrose constructed [9] a smooth region which is not illuminable from some (but not all) points, see Fig 1.

Mathematics Subject Classification 37D50

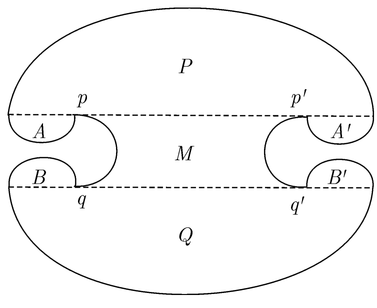


Fig. 1. The upper and lower curves are half-ellipses with foci of p, p' and q, q'. Using the fact that an ellipse reflects any ray passing between the foci back between the foci, it is easy to see that rays originating in regions P or M can never reach regions B and B'

Other authors modified this example to a smooth region not illuminable from any point. Thus, both questions were answered negatively for *smooth* regions. In 1978, J. Rauch gave [10] an example of a smooth region not illuminable from any finite set of points.

On the other hand, the solution for *polygonal* regions was not forthcoming and no significant progress appeared in the literature during for over 40 years. The nature of these problems, being easily stated and easily understood together with their apparent intractability had an obvious appeal. Thus, they started appearing on various lists of unsolved problems. V. Klee's paper, 1969 [6], seems to be the first published version. This was followed by a survey article of V. Klee and R. Guy, 1971 [5], and then by Klee's list of the ten most appealing unsolved problems in plane geometry in which the illumination problem was his fifth (see [7]). In 1991, two books [8] and [3] of unsolved problems have been published both of which give excellent discussions of the two illumination problems.

The problem **(1)** for polygons has been remained open for over forty years and was solved **negatively** by G. W. Tokarsky in 1995 in his article [13]. Tokarsky constructed examples of non-illuminable polygonal rooms all of which are *non-convex* polygons. The simplest example of such a room is shown in Fig. 2: there is no ray from the source at point A_0 that comes to point A_1. The proof of this statement is based on the following simple lemma (Lemma 3.1. in [13]): *In an isosceles right triangle ABC with right angle C, there no billiard trajectories starting at the vertex A and coming back to A.* (The proof of this lemma is, in turn, based on the evident fact that the straight line segment on a plane with endpoints at $(0,0)$ and $(2m,2n)$ passes through the point (m,n). Therefore, any ray starting at the vertex A and coming back to the same vertex must first pass through one of the vertices B or C of $\triangle ABC$, hence before hitting vertex A it must be terminated at B or C — a contradiction.)

The second problem, **(2)**, is still open for polygonal regions.

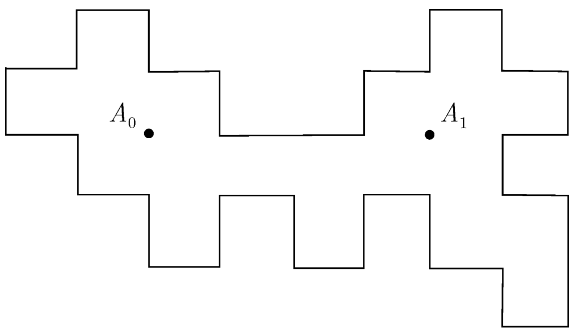


Fig. 2. There do not exist any pool shots from A_0 to A_1 on this polygonal non-convex table

The negative results described above show that the illumination problem should be modified. In the present article, we modify the problem as follows: **(a)** an arbitrary 2-dimensional region Q

is replaced by a **convex** region, and **(b)** the source S is replaced by a "cone" of light or a **search light** — the set of all rays inside a very small angle that are issued from the vertex of this angle.

Note that the tiny "cone" has uncountable set of rays, so the new "convex illumination problem" is close to the initial one with the source of rays issued in all directions (or, equivalently, with the 360° search light inside a non-convex polygon). On the other hand, if table Q has corners, e.g. Q is a convex polygon, a countable set of light rays issued from the search light will eventually disappear, and, perhaps because of that, Q will not be illuminable. It turned out that this problem is not trivial even for the case of a square (the result on square had been presented in [**?**]). The question about illumination of a polygon by the search light is the first question in the literature about polygonal billiards the answer to which depends on convexity of the table Q.

It is also natural to modify this problem for regions close to convex polygons: for dispersing (Sinai) and semidispersing billiards whose boundaries contain smooth curves concave inwards, while the other sides are straight line segments.

The present article is devoted to a complete investigation of the search light illumination problem for the following regions Q: **(I)** circles and ellipses; **(II)** squares; and **(III)** dispersing and semidispersing billiards. The results of investigations are the following.

In cases **(II)** and **(III)**, the billiard table Q is illuminated entirely, independently of the position of the search light S (inside Q or on ∂Q). (See Theorems 3.1 and 4.2.) In case **(I)**, three possibilities can occur: the illuminated region is either the entire ellipse, or the entire annulus between ∂Q and a confocal ellipse, or a region between two confocal hyperbolas (See Theorem 3.3.)

The entire illumination of the tables in cases **(II)** and **(III)** leads to the following conjecture for convex polygonal billiards:

Conjecture. *For any convex polygon Q, any search light S, and any position of the search light on the table Q, the whole table will be illuminated entirely.*

At the end of section 4, we show that if $Q =$ Weyl's chamber (i.e., an equilateral triangle, an isosceles right triangle, a right triangle with angles 30° and 60°, or a regular hexagon), then Q is illuminated entirely except, perhaps, a *finite* number of points (Theorem 3.6).

We use the following notation throughout the article:

Q is our billiard table, a bounded domain in $\mathbb{R}^2$ or $\mathbb{T}^2$ with piecewise smooth boundary.

$\mathcal{M} = Q \times S^1$ is phase space of the billiard flow Φ^t with smooth invariant measure μ.

M is the standard cross-section, consisting of unit vectors attached to ∂Q and pointing inward Q. The first return map $T : M \to M$ preserves a smooth measure ν.

On M, we use coordinates (r, φ), where r is the arclength parameter along ∂Q, and $\varphi \in [-\pi/2, \pi/2]$ the angle with the inward normal to ∂Q.

The search light is described as follows. $A \in Q$ is the source of light (the vertex of the search light beam), u is a unit vector with foot point at A called the *search light (unit,* or *bisector) vector.* We consider all the light rays coming out of A and making angles with the vector u less than some small fixed $\beta > 0$. The union of all those light rays (taken from A to the point of intersection with ∂Q) is called the beam of light. The triple (A, u, β) completely specifies the beam of light, see Fig. 3. Note that we do not include into the beam its edges (the rays making the angle $= \beta$ with the vector u), so the beam is open.

$L \subset Q$ is the part of the billiard table illuminated by the search light. That is, L consists of points $B \in Q$ such that there is a finite billiard trajectory starting at A with an initial velocity vector making angle $< \beta$ with u (i.e. belonging to our beam (A, u, β)) and hitting B, possibly after some reflections at ∂Q but before it hits any corner point of the table Q.

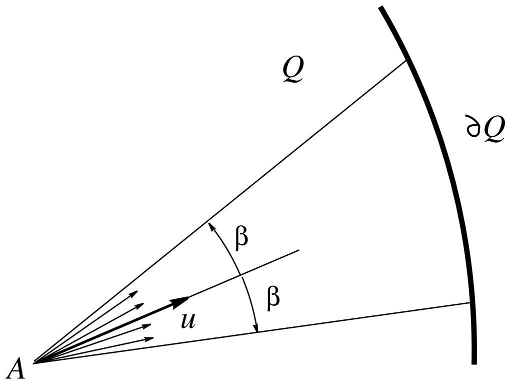


Fig. 3. Beam of light on a billiard table

2. Search light in a circle and an ellipse

Let Q be a circle of radius $R > 0$ with center O. For each $d \in (0, R)$ we will denote by $K(d) \subset Q$ the closed disk of radius d centered at O. There are three possibilities:
Case A: $A = O$. If so, each light ray emanating from A runs back and forth along a diameter of the circle Q. In that case L clearly consists of two sectors in Q: one made by the bundle (A, u, β) itself, and the other made by its symmetric image across the center O. No other parts of Q will be illuminated (Fig. 4 a).

Now assume that $A \neq O$ and denote by ℓ the line passing through O and A. Let γ denote the (non-obtuse) angle between the line ℓ and the vector u. There are two distinct cases:
Case B: $\gamma < \beta$. Then one of the light rays emanating from A lies on the line ℓ, and so the center O will be illuminated (see Fig. 4 b);

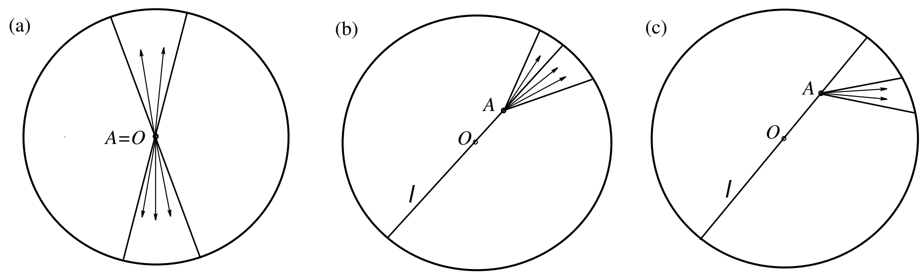


Fig. 4. Three cases for the location of the search light on a circular table

Case C: $\gamma \geqslant \beta$. Note that now the search light (A, u, β) lies on one side of the line ℓ (see Fig. 4 c). In that case all the light rays of the beam and their extensions beyond the point A (to the other side of ℓ) make two opposite sectors with vertex A, see Fig. 5. Denote by $d_0 > 0$ the distance from O to the union of these sectors. Note that if $\gamma = \beta$, then $d_0 = 0$.

Theorem 2.1. *In Case B, we have $L = Q$, i.e. the entire circle will be illuminated. In Case C, we have $L = Q \setminus K(d_0)$, i.e. the illuminated region is a ring. (In particular, if $\gamma = \beta$, then $L = Q \setminus \{O\}$).*

Proof.

Billiard trajectories in a circle have the following simple properties:
(a) every trajectory gets reflected off ∂Q with the same angle of reflection, i.e. φ is preserved by the map T;

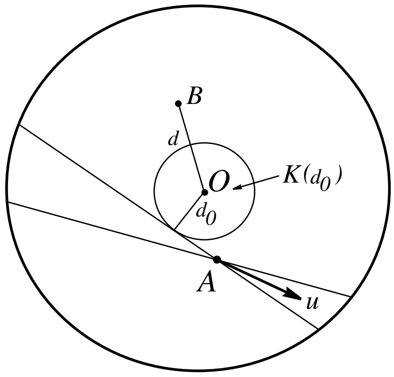


Fig. 5. Beam of light in a circle: Case C

(b) the distance d from every segment of a trajectory to the center O is constant, i.e. every link of the trajectory is tangent to the circle of radius d centered at O. That circle is called a *caustic*.
The property (b) immediately implies that no point of $K(d_0)$ can be illuminated in Case B. To prove the theorem it now suffices to show the following:

Lemma 2.2. *Let $B \in Q \setminus K(d_0)$. Then $B \in L$.*

Proof.
Note that $d := \mathrm{dist}(B, O) > d_0$. Then there is a narrow subbundle of light rays in our beam such that the distances from O to all of the rays in that subbundle (or their continuation beyond the point A) are less than d. When this subbundle falls on ∂Q and gets reflected, the outgoing velocity vectors will make a continuous curve $W \subset M$. The curve W cannot entirely lie on any one line $\varphi = \mathrm{const}$ in the phase space M, hence it intersects the lines $\varphi = c$ for all c in some narrow interval (c_1, c_2), see Fig. 6.

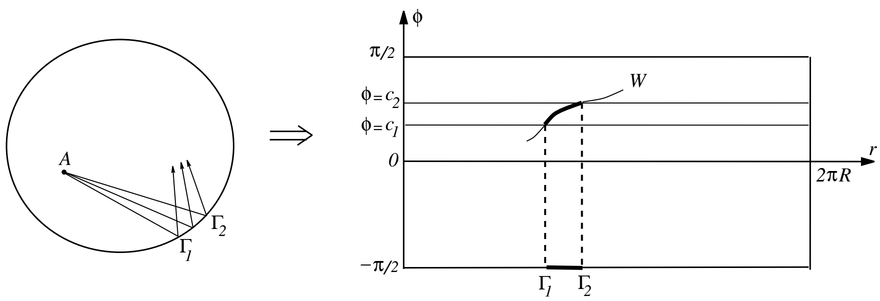


Fig. 6. The arc $\Gamma_1\Gamma_2 \subset \partial Q$, the outgoing velocity vectors, and the image W on the cross-section M

Now consider the images $T^n W$ as $n \to \infty$. Note that M is a cylinder in the coordinates (r, φ), since r is a cyclic coordinate with period $2\pi R$. Every line $\varphi = c$ is a circle on that cylinder, which is invariant under T, and the coordinate r on it changes by the rule

$$r \mapsto r + R(\pi - 2c) \pmod{2\pi R}$$

see Fig. 7. Hence, the line $\varphi = c$ experiences a rigid rotation under T through the angle that continuously and monotonically depends on c. Therefore, the curve $T^n W$, as n grows, will get stretched in the r direction and eventually "wrap around the cylinder" M. In fact, the curve $T^n W$ will tend to fill the narrow cylinder $\{c_1 < \varphi < c_2\} \subset M$ more and more densely as $n \to \infty$.

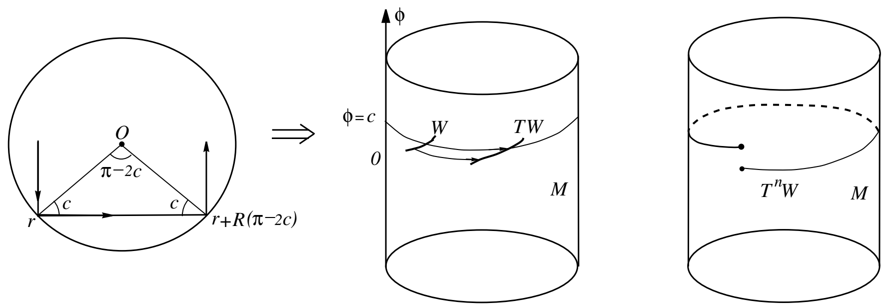


Fig. 7. Rotation of the line $\varphi = c$ under the map T

It is enough for us to note that for sufficiently large n, there is a continuous subcurve $W' \subset T^n W$ along which the r-coordinate changes from 0 to $2\pi R$. The outgoing velocity vectors of the curve W' make a continuous family of light rays, one of which necessarily hits the point B, as it is clear from Fig. 8. Lemma is proved. ■

In fact, the first statement of Theorem 2.1 also follows from Lemma 2.2. ■

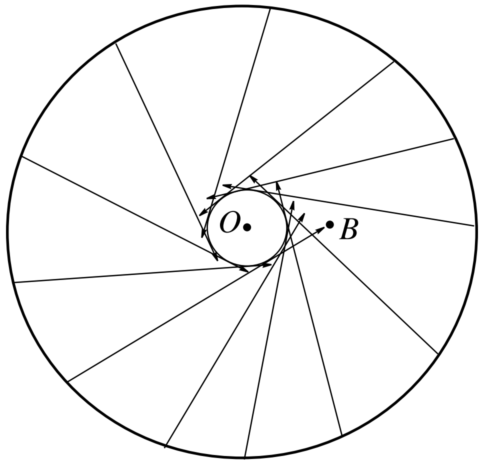


Fig. 8. Light rays of the curve W'

The search light problem in an ellipse is very similar to that in a circle. In an elliptic billiard table, every link of any trajectory (except those passing through the foci) is tangent to a caustic. However, there are two types of caustics: confocal ellipses and confocal hyperbolas. A detailed description of billiard trajectories in an ellipse and their caustics can be found in [12].

Fig. 9 illustrates the structure of the cylinder M. It is foliated by T-invariant curves of two types. Those curves whose trajectories have a hyperbolic caustic make a figure ∞ (shaded "eye glasses"). Each invariant curve of this type consists of two ovals, one in each shaded "eye". The curves that correspond to trajectories with an elliptic caustic fill the regions above and below the figure ∞. The boundary of the figure ∞ corresponds to trajectories passing through the foci of the given ellipse. The map T (T^2) acts as a rotation on each invariant curve outside (inside) of the figure ∞. Moreover, the angle of rotation depends continuously and monotonically on the curve in both regions.

Consider a search light (A, u, β) in an ellipse Q. Continue every light ray in this beam until it crosses the major axis of the ellipse for the first time (this might happen after a few reflections off the boundary). The points of intersections of our rays with the major axis of the ellipse will make an interval on that axis. Call that interval I, see Fig. 10.

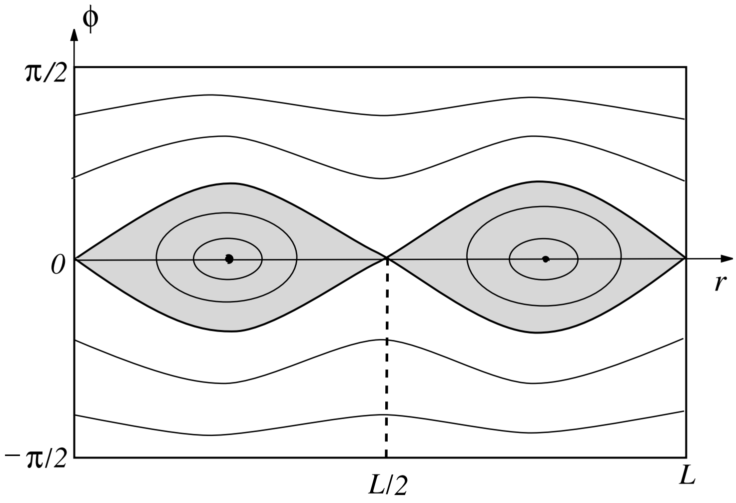


Fig. 9. Invariant curves on the cylinder M for an elliptical billiard

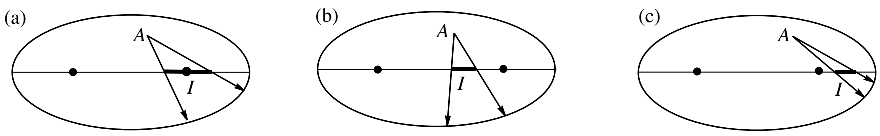


Fig. 10. Three cases for the location of the search light on an elliptical table

Theorem 2.3. *If the interval I covers a focus, then $L = Q$, i.e. the entire ellipse will be illuminated. If I lies between the two foci, then the illuminated region L lies between two confocal hyperbolas. If I lies on the major axis outside the interfocal segment, then the illuminated region L lies between the given ellipse and a confocal ellipse.*

The two cases where L is not the entire ellipse Q are illustrated on Fig. 11.

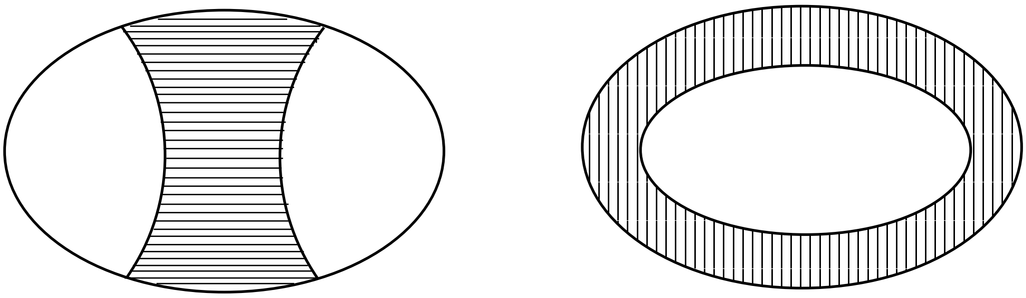


Fig. 11. Illuminated regions in an ellipse

Proof.

As in the case of a circular table, consider the light rays passing through the interval I and their first reflection off ∂Q. The outgoing velocity vectors make a curve W on the phase cylinder M.

Now there are three cases: **(a)** W intersects the boundary of the figure ∞ (i.e., one of the rays passes through a focus), **(b)** W lies inside the figure ∞, and **(c)** W lies outside of the figure ∞.

The cases **(b)** and **(c)** are similar to the case C in Theorem 2.1. The image T^nW, for large n, will be arbitrarily long and we can argue as before. The case (a) is similar to the case B in Theorem 2.1. We leave the details to the reader. ■

3. Square billiard tables

In this section, we consider illumination of square billiard tables. For some positions of the search light $S = (A, u, \beta)$ inside square Q, it is almost evident that the whole square will be illuminated entirely. For example, if vertex A of the search light is located at the center of the square Q and the angle bisector, u, of the beam of light is orthogonal to a side of the square, then Q is illuminated entirely. Indeed, using the standard procedure of unfolding of a beam of light (or, for the sake of simplicity, the search light) on the square lattice $\mathbf{Z}^2$ along the line λu, $\lambda \in \mathbb{R}^+$, we can easily see that these unfolded images contain an entirely illuminated square cell $Q' \in \mathbf{Z}^2$ which is an image of the table Q under unfolding (see Fig. 12. Since Q' is directly illuminated by the beam of light, so is the table Q.

The same happens whenever vector u is orthogonal to a side of the table Q, regardless of the A's position inside the table Q.

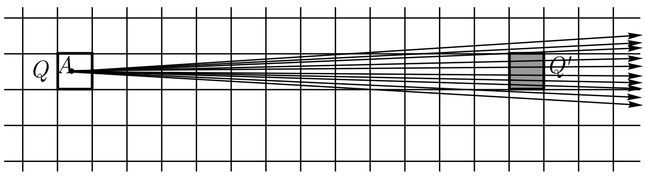


Fig. 12. A particular case of illuminating a square table Q

However, if vector u is not orthogonal to a side of Q, there is no a square cell in the lattice $\mathbf{Z}^2$ that is directly illuminated by the unfolded image of the search light. Figure 13 shows how the countable set $\{P\}$ of lattice points ("nodes") inside the unfolded image of the search light cast shadows — the "black" rays started at points P in the direction of $\overrightarrow{AP}$. These shadows consist of the points that are non-illuminable directly by the search light. The "black" rays form, in turn, "black" segments inside the lattice squares belonging to or intersecting with the angle A, namely, the images $\{Q'\}$ of the initial square Q under repeated reflections in Q's sides. Those points in squares Q' are not illuminated from the source A directly, i.e. inside the angle A. Coming back to the initial square Q, we can say that the "black" points (points on the "black" shadow segments) could have been illuminated by the light rays issued from A if they were not absorbed at the corners of square Q. (The absorbtion creates "black" rays in the unfolded images.) Nevertheless, it could happen that some other, non-directed (and hence non-absorbed), lights issued from A will illuminate all the points on black segments, in which case the table Q would be illuminated entirely.

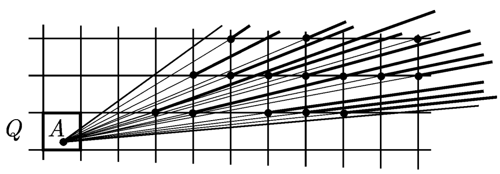


Fig. 13. The shadows from lattice nodes are not illuminated directly from the vertex A

Another difficulty comes when the search light is situated at the boundary of the table, or, especially, at one of its vertices. The aim of this section is to prove that regardless of the search light position, the whole square is going to be illuminated entirely.

Theorem 3.1. *If Q is a square billiard table, then $L = Q$ for any search light $S = (A, u, \beta)$.*

Our proof of the Theorem 3.1 is based on special properties of periodic billiard trajectories in the square table.

We formulate and prove the following three lemmas about periodic billiard trajectories in a square. For the sake of simplicity, we will think of the table Q as of a unit (i.e. 1×1) square $\{(x, y)|\ x, y \in [0, 1]\}$ and, at the same time, think of the standard unit square lattice $\mathbf{Z}^2$ as of the union of all possible unfolded images of the square Q (under all possible reflections in Q's sides). We will also consider a billiard trajectory γ inside Q together with its forward (sometimes backward) unfolded images, the ray $\vec{r}(\gamma)$ on the plane $\mathbb{R}^2$; the square cells of $\mathbf{Z}^2$ that ray $\vec{r}(\gamma)$ intersects, form the *corridor* for the unfolding of γ. Likewise, we associate each ray $\vec{r}(\gamma)$ on the plane with a billiard trajectory γ for which $\vec{r}(\gamma)$ is its unfolding on the plane.

We start with the very famous and standard lemma written in many textbooks on dynamical systems; we omit its proof.

Lemma 3.2. *Let γ be a billiard trajectory started at a non-corner point A of the square table Q, and φ be the angle formed by the direction of the trajectory and the positive x-axis. Then γ is periodic if and only if* $\tan\varphi$ *is a rational number and the ray $\vec{r}(\gamma)$ does not pass through a $\mathbf{Z}^2$ node. Moreover, each nonperiodic billiard trajectory γ fills the table Q everywhere densely.*

Next lemma concerns periodic billiard trajectories inside the search light S.

Lemma 3.3. *Let vertex A of the search light S be any point either inside Q or on a side of Q but not a Q's vertex. Then, for any value of the search light angle β and for any direction of the search light unit vector u, at least one of the search light rays issued from point A is a periodic billiard trajectory, γ, in Q. In other words, there is a ray $\vec{r}(\gamma)$ with the vertex A that lies entirely inside the unfolded image of the search light (i.e. inside $\angle A$) and does not pass through any of the $\mathbf{Z}^2$ nodes. (See Fig. 14.)*

Proof.

<u>**Case 1.**</u> Suppose that A is not a rational point on the plane, i.e. at least one of A's coordinates is an irrational number. Let us pick two rays $\vec{r}(\gamma_1)$ and $\vec{r}(\gamma_2)$ inside $\angle A$ both with <u>rational</u> slopes, $m_1 > m_2$. Then at least one of the rays does not contain a node of $\mathbf{Z}^2$, otherwise (if both rays contain nodes of $\mathbf{Z}^2$) point A would have been a rational point (with two rational coordinates) as the intersection point of two rational lines on the plane; this is a contradiction with our supposition. Therefore, at least one of the two billiard trajectories, γ_1 or γ_2, is periodic by Lemma 3.2.

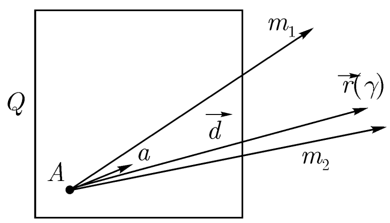


Fig. 14. Ray $\vec{r}(\gamma) = A + t \cdot \vec{d}$ does not pass through a lattice point and is the unfolded image of a periodic billiard trajectory γ

<u>**Case 2.**</u> The second case happens when point A is a *rational* point on the plane — the point of intersection of two rays $\vec{r}(\gamma_1)$ and $\vec{r}(\gamma_2)$ inside $\angle A$ both having <u>rational</u> slopes $m_1 > m_2$ and both containing lattice nodes. In this case, none of the trajectories, γ_1 or γ_1, is a billiard trajectory because each one terminates at some of the Q's corners. Let $A = (a_1, a_2) = \left(\frac{k}{\ell}, \frac{m}{n}\right)$, where k, ℓ, m, and n are positive integers, and $\frac{k}{\ell} > \frac{m}{n}$ are both irreducible fractions.

Let us fix a big prime number p such that $p \cdot \frac{m}{n} \notin \mathbf{Z}$ and consider any positive integer $q \in \mathbf{Z}$ satisfying $m_2 < q \cdot \frac{\ell}{p} < m_2$. Then vector $\vec{d} = \left(1,\ \frac{q\ell}{p}\right)$ is clearly directed inside the angle $\angle A$, and, hence, the line

$$\vec{r}(\gamma) = A + t \cdot \vec{d} = (a_1, a_2) + t \cdot (1,\ q \cdot \frac{\ell}{p}) = \left(\frac{k}{\ell} + t,\ \frac{m}{n} + \frac{ql}{p}t\right)$$

lies entirely inside of $\angle A$. (Parameter t runs throughout the positive semi-line $\mathbb{R}^+$).

Claim: *The ray $\overrightarrow{r}(\gamma) = A + t \cdot \overrightarrow{d}$ does not contain a lattice point.*

Proof of the Claim. Assuming the contrary, denote by (a, b) a lattice point that belongs to the ray $\overrightarrow{r}(\gamma)$. Then

$$\begin{cases} \frac{k}{\ell} + t = a \in \mathbf{Z}^2, \\ \frac{m}{n} + \frac{ql}{p} t = b \in \mathbf{Z}^2. \end{cases}$$

Then, making the substitution $t = a - \frac{k}{\ell}$, we can easily obtain that

$$p \cdot \frac{m}{n} = bp - q(\ell a - k).$$

We get a contradiction because the left hand side of the equality is not an integer by our assumption, whereas the right hand side is clearly an integer. The Claim is proven. ∎

Thus, the ray $\overrightarrow{r}(\gamma)$ does not pass through a $\mathbf{Z}^2$ node and hence can be thought of as an unfolded image of a billiard trajectory γ. Since the slope of the ray $\overrightarrow{r}(\gamma)$ is a rational number, trajectory γ is periodic inside the square Q. The lemma is proven.

The proof of Lemma 3.3 does not change for the the following its generalization:

Lemma 3.4. *Let $A \in \mathbb{R}^2$ be a rational point different from the origin O. Then the set of rational lines that pass through A and do not pass through any node of the lattice $\mathbf{Z}^2$, is everywhere dense on the circle S^1 of all directions.*

Remark. The statement of Lemma 3.3 does not hold if vertex A of the search light is situated at a vertex of square Q, say, at the origin $O = (0, 0)$. In this case, any ray with rational slope issued from A contains infinitely many nodes of $\mathbf{Z}^2$. Indeed, any ray with a rational slope coming out of the origin O contains a rational point $\left(\frac{m}{n}, \frac{p}{q}\right)$, and hence contains the node $(mq,\ np)$ of the lattice $\mathbf{Z}^2$.

Next lemma will play the most important role in the proof of Theorem 3.1. Before formulating this lemma, we remind the basic notions assigned to a billiard trajectory: "the code of a trajectory", "equivalent trajectories", "the family of equivalent trajectories", and "a generalized diagonal".

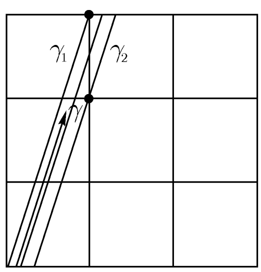
Q 

Fig. 15. The strip $[\gamma]$ of equivalent periodic billiard trajectories bounded by the generalized diagonals γ_1 and γ_2

The *code* of a billiard trajectory γ is the sequence of Q's sides that the billiard point hits moving along the trajectory γ. Trajectory γ is periodic if and only if its code is a periodic sequence. Two trajectories are *equivalent* if they have the same code. Two equivalent periodic billiard trajectories have the same length and are represented in their unfolded images by two congruent parallel segments. A *generalized diagonal* (or a *billiard diagonal*) is a finite piece of a billiard trajectory outgoing from a Q's vertex and incoming to a Q's vertex (it is not a real billiard trajectory because it terminates in corners of the table). Consider the family $[\gamma]$ of all periodic billiard trajectories γ' equivalent to γ. In the unfolded images, they form a finite-width open strip of parallel congruent segments bounded by the unfolded images of two generalized diagonals of the same length, γ_1 and γ_2 (see Fig. 15). Inside Q, the strip $[\gamma]$ reflects off the Q's sides as the bundle of parallel light rays and form the *corridor* of trajectories equivalent to γ which we also call $[\gamma]$. The closure of the open strip (corridor) $[\gamma]$ is the closed strip (corridor) with the boundary $\gamma_1 \cup \gamma_2$.

Lemma 3.5. *Let γ be a periodic billiard trajectory in the square Q emanated from vertex A inside the search light $S = (A, u, \beta)$. Then the closure $\overline{[\gamma]}$ of the corridor $[\gamma]$ of equivalent to γ periodic trajectories covers the whole square Q entirely.*

Proof.

Each billiard trajectory from the corridor $[\gamma]$ consists of segments parallel to exactly two directions of the corridor (see Fig. 16). We call the segments along the first direction "the segments of type (1)", and those parallel to the other direction "the segments of type (2)".

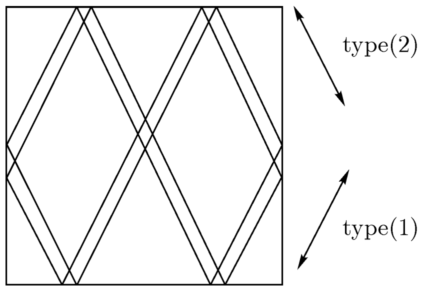


Fig. 16. Two types of segments of a periodic trajectory equivalent to the trajectory γ

Consider only the part of the corridor formed by the segments of type (1), and call it "part (1) of the corridor".

Claim: *Part (1) of the corridor covers the square Q entirely.*

Proof of the Claim. Part (1) of the corridor consists of several parallel corridor strips with possibly some gaps between them (see Fig. 17 a). We will prove that there are actually no gaps between the strips, or all the gaps have zero width. (We will call a gap with non-zero width "a non-zero gap".)

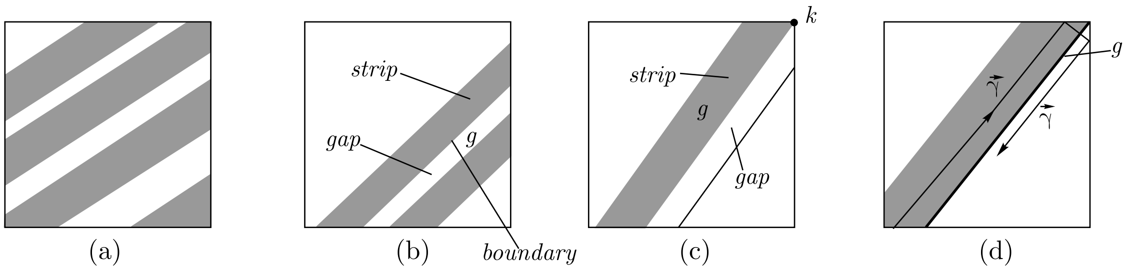


Fig. 17. The strip of the corridor $\overline{[\gamma]}$ and the gap with the same boundary g will eventually overlap after two reflections at the corners of the square K

Suppose, on the contrary, that there is a non-zero gap that lies next to some corridor strip. Denote by g the common boundary (segment) of this gap and this strip (Fig. 17 b). Let us prolong both the strip and the gap inside the square Q (making billiard reflections off the corresponding Q's sides) until the boundary g hits a square's vertex — point K on Figure 17 c.

Then consider the reflection of the strip near the vertex K. After two reflections of the strip from the two adjacent sides of Q with the common vertex K, it will overlap the gap, at least partially (look at the light ray $\tilde{\gamma}$ on Fig. 17 d that after two reflections moves inside the gap). We get a contradiction.

So, both the claim and the lemma are proven. **Q.E.D.** ■

Now we are ready to start proving Theorem 3.1.

Proof of Theorem3.1. We consider two cases for the position of the search light's vertex A: A is not a corner (vertex) of square Q, and A is a Q's corner.

Case 1: *A is not a corner (vertex) of square Q.*

Let's P be any point in the square.

Claim 1: *Point P is illuminated by the search light S.*

Proof of Claim 1. Consider a periodic trajectory γ that comes out from vertex A inside the search light S (i.e. inside $\angle A$); such a trajectory exists according to Lemma 3.3. As before, denote by $[\gamma]$

the corridor of equivalent to γ periodic trajectories inside Q (they are not rays of light issued from the source A!). All the trajectories from corridor $[\gamma]$ avoid the Q's corners (vertices) and, according to Lemma 4.5, $\overline{[\gamma]} = Q$. Hence, either **(1)** $P \in [\gamma]$ or **(2)** $P \in \partial\overline{[\gamma]}$. We want to prove that there is a light ray from A that passes through point P after several reflections from the Q's sides.

Let us unfold the corridor $[\gamma]$ — we get an infinite strip of rays parallel to ray $\vec{r}(\gamma)$; the boundary of the strip is the union of the unfolded images of two generalized diagonals parallel to $\vec{r}(\gamma)$. This strip partially overlaps with the search light, or angle $\angle A$, but the infinite part of the strip (outside a big circle centered at A) lies inside $\angle A$ entirely (Fig. 18).

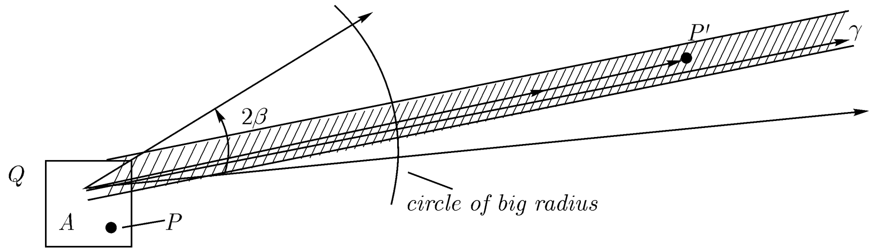


Fig. 18. A far image, P', of point P belongs to the unfolded images of the strip $[\gamma]$ and is illuminated by the light ray $\overrightarrow{AP'}$

The unfolded images of $[\gamma]$ contain infinitely many images of point P under multiple reflections of point P in Q's sides, because $\overline{[\gamma]} = Q$ by Lemma 3.5. Consequently, infinitely many images of point P belong to $\angle A$. Let P' be one such image of point P located far away from point A. Connect points A and P' by a straight line segment. Then the light ray $\overrightarrow{AP'}$ illuminates point P', and consequently, its preimage, P, is illuminated by the same search light ray emanated from A in the direction $\overrightarrow{AP'}$. Claim 1 is completely proven.

Case 2: *A is located at a vertex of square Q.*

In this case, Lemma 3.3 fails: there is no ray with a rational slope coming out of A that does not contain a lattice point on it (see the remark after Lemma 3.4). It means that all rational rays issued from vertex A of the search light S will disappear in table's corners (vertices). So, Lemma 3.3 is not applicable for our aim directly: we cannot construct a periodic trajectory inside Q and hence cannot construct a corridor of equivalent periodic trajectories that would cover the whole square as it took place in Lemma 3.4. However, Lemma 3.3 can be used indirectly for constructing a different corridor of periodic orbits inside Q.

Let us draw a *rational* ray $\vec{r}$ from point O. This ray contains lattice points on it, and let V be the closest to O lattice point; thus, the straight line segment OV does not contain a node of $\mathbf{Z}^2$ on it. We may choose as $\vec{r}$ the ray for which the length of the segment OV as big as we wish; we will not specify this in the further text. Consider in table Q a periodic billiard trajectory γ with the same slope as the drawn ray $\vec{r}$ (such periodic orbit exists according to Lemma 3.2), and consider, in addition, the widest possible corridor $[\gamma]$ of periodic orbits equivalent to the trajectory γ together with its unfolding (denoting also by $[\gamma]$) — the strip of unfolded images of all periodic billiard trajectories equivalent to γ. The boundary of this corridor contains segment OV (because OV is an unfolded image of the generalized diagonal in the direction of $\vec{r}$). Note that, as before, an infinite part of the strip $[\gamma]$ (outside a circle of big radius with center O) lies inside the search light, i.e. inside $\angle A$.

Consider any point P on the table Q and its images under all possible reflections in Q's sides (one image per each square of $\mathbf{Z}^2$). By Lemma 4.3, the closure of the corridor $[\gamma]$ covers the whole table Q entirely. Consequently, one of the P's images under all possible reflections, say P', must belong to the strip $\overline{[\gamma]}$. We pick an image P' that belongs to both the strip and the angle $\angle A$ (in other words, P' is located very far from point O). There are three possibilities for point P' to locate in the strip $\overline{[\gamma]}$.

Possibility 1: *P' is located inside the strip* $[\gamma]$ (Fig. 19).

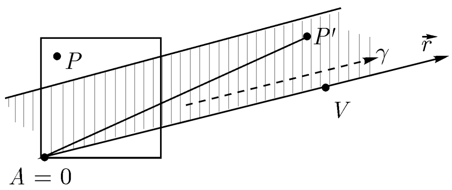


Fig. 19. Illumination of a point P' located inside the strip $[\gamma]$

Then segment OP' lies inside the strip $[\gamma]$ and does not contain a node of $\mathbf{Z}^2$ on it. Therefore, point P' is illuminated by the light ray $\overrightarrow{OP'}$. Consequently, point P on the table Q is also illuminated by the same light ray $\overrightarrow{OP'}$.

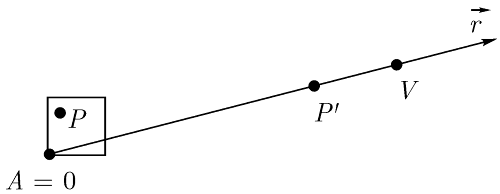


Fig. 20. Direct illumination of a point $P' \in OV$

Possibility 2: *P' is an interior point of segment OV* (Fig. 20).

Then, clearly, point P' is illuminated directly by the light ray $\overrightarrow{r}$ emanated from vertex O of the search light S. Therefore, point P on the table is also illuminated by the light ray $\overrightarrow{r}$ directly.

Possibility 3: *Lattice point V is an interior point of segment OP'* (Fig. 21).

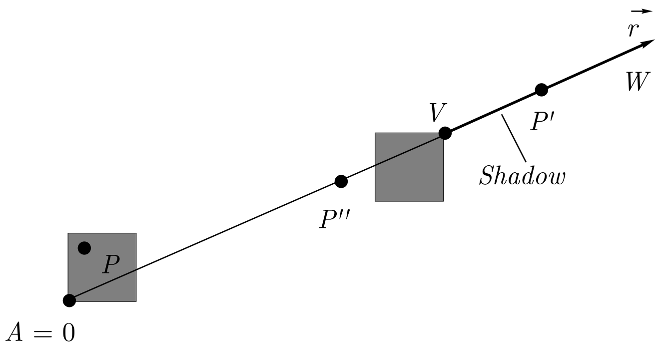


Fig. 21. Indirect illumination of a point $P' \in OV$

In this case, point P' belongs to the segment VW, where W is the next node of $\mathbf{Z}^2$ after node V on the ray $\overrightarrow{r}$, i.e. $OV = VW$. Although point P' belongs to the shadow cast by vertex V and therefore is not illuminated by the light ray $\overrightarrow{r}$, the equivalent to P' under multiple reflections in Q's sides point $P'' \in OV$, for which $VP'' = VP'$, is illuminated by the light ray $\overrightarrow{OV}$ (the segment OV does not contain lattice points on it!). Therefore, point P on the table Q, as a preimage of point P'', will also be illuminated in this case.

We considered all cases for the location of the search light and proved in each case that arbitrary point $P \in Q$ is illuminated by a light ray issued from the source A of the search light S. Thus, Theorem 3.1 is completely proven. **Q.E.D.** ■

As a conclusion of this section, we give a very simple proof of the following theorem on illumination of a Weyl's chamber: a polygon Q, multiple reflections of which in its sides form a lattice on the plane $\mathbb{R}^2$.

Theorem 3.6. *If Q is a Weyl's chamber: a rectangle, or an equilateral triangle, or an isosceles right triangle, or a right triangle with angles $30°$ and $60°$, or a regular hexagon, then for any position of the search light $S = (A, u, \beta)$, table Q will be illuminable except a finite number of points.*

Proof.

Let us reflect Q in its side to get a regular lattice on the plane (rectangular, square, triangular, or hexagonal). Since the sides of the angle $\angle A$ diverge to infinity, we can find (very far from the vertex A, if the value of β is very small) two copies of polygon Q inside $\angle A$ with a common side. See Fig. 22, where for the sake of simplicity, Q is taken to be a rectangle.

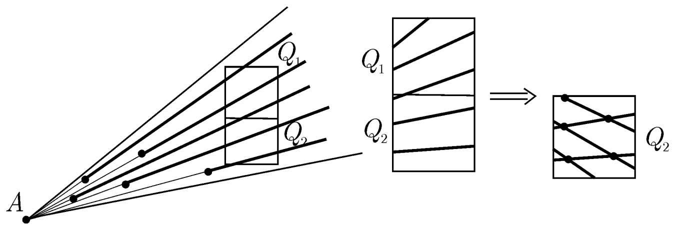


Fig. 22. (a) Two copies of the table Q with a common side contain a finite number of divergent shadow segments cast by lattice's nodes; (b) Reflecting in the common side gives only a finite number of non-illuminable points on table Q

Denote those copies of the table by Q_1 and Q_2. Let us bend (reflect) Q_2 with respect to the common side of Q_1 and Q_2. Since the shadow segments on Q_1 and Q_2 are divergent (all of them are issued from the same point A), after reflection, the shadow segments from Q_2 intersect with the shadow segments on Q_1 at a finite number of points.

It is clear that all the points on the shadow segments except the finite number of intersection points of the two families of segments are illuminated. Therefore, all the points on the table Q except, perhaps, the mentioned intersection points, are illuminated. The theorem is proven. ■

Note. One can decrease the number of non-illuminable points by using the following operation. Consider in the proof of Theorem 4.6 not only two copies of adjacent polygons inside $\angle A$, but a connected chain of mutually adjacent copies of the table Q; this can be done if these polygons are far away from vertex A. Reflecting one polygon after another with respect to their common sides, we will "kill" the points which do not fall onto the shadow segment: those points will be illuminated by some light rays issued from the source A. We also can choose an infinite chain of such polygons inside $\angle A$ and make infinitely many reflections in common sides, decreasing each time the number of intersection points. Suppose now that the polygon Q will not be illuminated entirely. This means that no matter what chain of polygons we take, the number of intersection points will not become smaller, i. e., it will be stabilized after some moment. It is very doubtful, and most likely, after some reflection, all of the non-illuminable points of some polygon Q_i will not fall onto a shadow segments of the polygon Q_{i+1}, meaning that all points of the table Q are illuminated.

4. Sinai billiard tables

It is interesting that the search light problem can be solved completely for the class of billiards most popular in modern statistical physics and, normally, most difficult to study: Sinai billiards.

A billiard table Q is called a *dispersing* (or *Sinai*) *billiard* if its boundary ∂Q is concave inward at all regular points (where a tangent vector exists) – see Fig. 23 a. A table Q is said to be *semidispersing* if its boundary ∂Q is either concave inward or flat at all regular points — see Fig. 23 b. In this case we also assume that

(a) the curved (i. e. concave) part of the boundary is not empty;

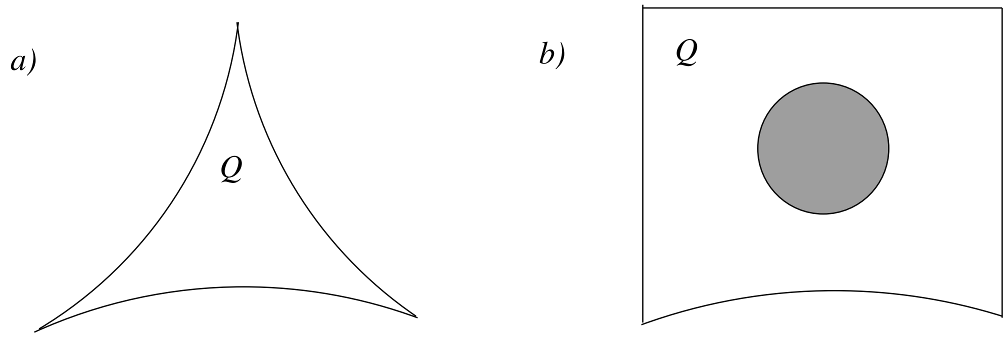


Fig. 23. A dispersing billiard table (a) and a semidispersing table (b)

(b) the trajectories that reflect at the flat parts of the boundary only (i. e., never hit any curved part of ∂Q in the future or in the past) make a set of zero measure and codimension at least two in the phase space M.
Note that it is not known at present time if **(b)** follows from **(a)** or not.

Theorem 4.1 ([11],[4]). *Let Q be a dispersing or semidispersing billiard table. Then the map $T\colon M \to M$ is hyperbolic, ergodic, mixing and Bernoulli.*

We refer the reader to [11], [1], [2] for definitions and proofs of this and related results.

Here we claim that

Theorem 4.2. *If Q is a dispersing or semidispersing billiard table, then $L = Q$ for any search light $S = (A, u, \beta)$.*

Proof.

Let $B \in Q$ be any point. Consider a narrow bundle of rays in Q that pass through B (i. e. "focus" at B), see Fig. 11. Continue these rays backwards until they intersect ∂Q, then the outgoing velocity vectors at the intersection points will make a curve $V \subset M$. Similarly, as the rays of our light beam emanating from A hit ∂Q and bounce off it, the outgoing velocity vectors will make a curve $W \subset M$, see Fig. 24. In the standard orientation of the r, φ coordinates [1], [2], expanding curves are increasing (have positive slopes), and contracting curves are decreasing (have negative slopes).

In the theory of dispersing and semidispersing billiards, the curves like V are called *stable* or *contracting*, and the curves like W are called *unstable* or *expanding*. It is quite clear that the expanding curve W will grow under the iterations of T, as shown on Fig. 24. Likewise, the contracting curve V will grow under the backward iterations of T (i. e. under T^{-n} as $n \to \infty$).

We note that if an expanding curve γ^e contracts under all the backward iterations of T (in particular, T^{-n} is continuous on γ^e for all $n \geqslant 1$), then γ^e is called an *asymptotically expanding curve* (or an *unstable manifold*). Likewise, if a contracting curve γ^c contracts under all the forward iterations of T, it is called an *asymptotically contracting curve* (or a *stable manifold*). It is known [1], [2], [11] that for almost every point $x \in M$ there are unique asymptotically expanding and contracting curves passing through x, and they are transversal to each other.

Our theorem now follows from

Lemma 4.3. *For any expanding curve $W \subset M$ and any contracting curve $V \subset M$ there is an $n \geqslant 1$ such that $V \cap T^n W \neq \emptyset$.*

Proof.

We only outline the argument, referring the reader to [2] for more details on the matter.

For almost every point of the curve V (and W), with respect to the length measure on that curve, there is an asymptotically expanding (resp., contracting) curve passing through that point,

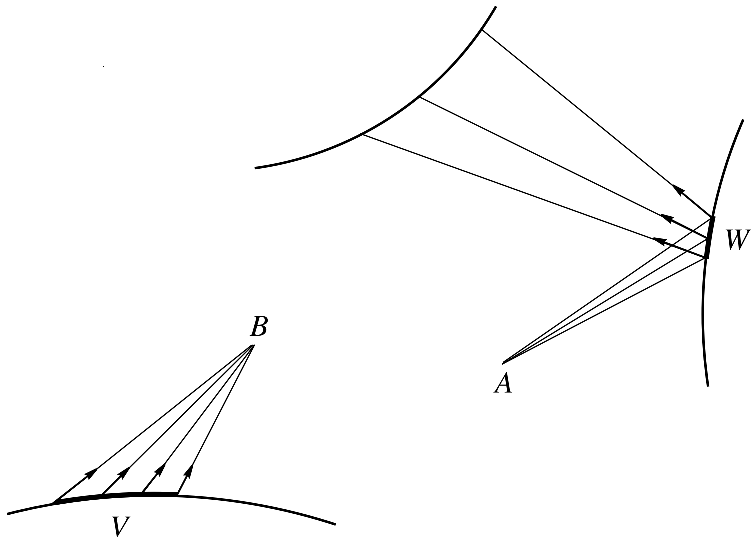


Fig. 24. Light rays coming out of the source A and light rays focusing at the target B

see [2]. With the help of these curves, one can construct two small parallelograms, Π_1 and Π_2, such that the curve W cuts through the middle of Π_1, and V cuts through the middle of Π_2, see Fig. 25. Such constructions are described in [2]. We recall that a parallelogram Π is made by the points of intersection of a family of asymptotically expanding curves $\mathcal{F}^e_\Pi$ with a family of asymptotically contracting curves $\mathcal{F}^c_\Pi$, provided every curve of $\mathcal{F}^e_\Pi$ intersects every curve of $\mathcal{F}^c_\Pi$ in exactly one point. It is usually required (and we do so) that $\mu(\Pi) > 0$.

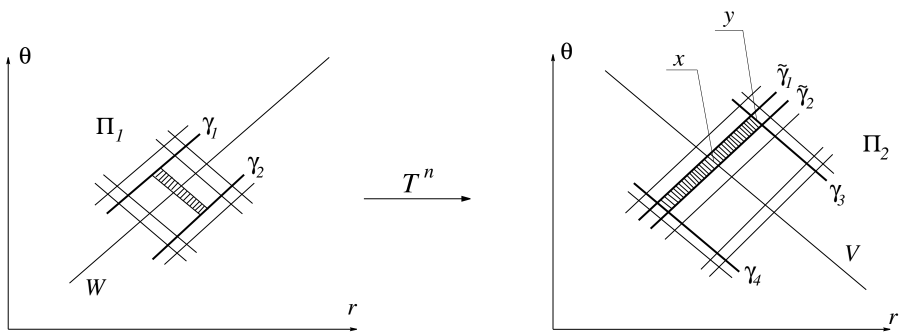


Fig. 25. Proof of Lemma 4.3. The shaded areas are the domains D and $T^{-n}D$

Now the points of intersection of the curve V with the curves of the family $\mathcal{F}^e_{\Pi_2}$ make a subset $R \subset V$ of a positive length measure on V (the latter follows from the fact $\mu(\Pi_2) > 0$ and the absolute continuity of asymptotic curves [2]). We pick any density point $x \in R$ and a narrow open interval $I \subset V$ containing x. Let $\tilde{\mathcal{F}}^e_{\Pi_2}$ be the subfamily of curves $\gamma \in \mathcal{F}^e_{\Pi_2}$ that intersect I. The points of intersection of the curves of $\tilde{\mathcal{F}}^e_{\Pi_2}$ with the curves of $\mathcal{F}^c_{\Pi_2}$ make a subparallelogram $\tilde\Pi_2 \subset \Pi_2$, and $\mu(\tilde\Pi_2) > 0$.

By the mixing property of T, we have $\mu(T^n\Pi_1 \cap \tilde\Pi_2) > 0$ for all sufficiently large n. If the interval I is small enough, there will be two curves $\gamma_1, \gamma_2 \in \mathcal{F}^e_{\Pi_1}$ lying on the opposite sides of W, such that $T^n\gamma_1$ and $T^n\gamma_2$ cover two curves of the family $\tilde{\mathcal{F}}^e_{\Pi_2}$, call them $\tilde\gamma_1$ and $\tilde\gamma_2$. Pick a point $y \in T^n\Pi_1 \cap \tilde\Pi_2$ and let $\gamma_3 \in \mathcal{F}^c_{\Pi_2}$ be the curve that passes through y. Since it lies on one side of V, we pick arbitrarily another curve $\gamma_4 \in \mathcal{F}^c_{\Pi_2}$ that lies on the opposite side of V, see Fig. 25.

By construction, the map T^{-n} is continuous of $\tilde{\gamma}_1$, $\tilde{\gamma}_2$ and on the part of γ_3 between $\tilde{\gamma}_1$ and $\tilde{\gamma}_2$. The singularities of the map T have a characteristic *continuation property* [1], [2], by which the map T^{-n} must be continuous on the entire domain D bounded by the curves $\tilde{\gamma}_1$, $\tilde{\gamma}_2$, γ_3, and γ_4. Loosely speaking, the map T^{-n} is singular along some expanding curves that can be continued in each direction up to ∂M. Hence, if T^{-n} is anywhere discontinuous inside D, the corresponding singularity curve stretches across D completely and must intersect at least two sides of D, which is impossible.

Now observe that the domain $T^{-n}D$ covers a piece of W that lies between γ_1 and γ_2. On the other hand, $T^{-n}V$ stretches across $T^{-n}D$ from γ_1 to γ_2, and hence intersects W. This proves the lemma. ■

Our argument is illustrated on Fig. 25. A similar argument was used and described in detail in [2].

Acknowledgement. We thank E. Gutkin for reading the manuscript and making many helpful remarks. N. Chernov was partially supported by NSF grant DMS-0098788.

References

[1] *L. A. Bunimovich, Ya. G. Sinai and N. I. Chernov.* Markov partitions for two-dimensional billiards. Russ. Math. Surv. V. 45. 1990. P. 105–152.

[2] *L. A. Bunimovich, Ya. G. Sinai and N. I. Chernov.* Statistical properties of two-dimensional hyperbolic billiards. Russ. Math. Surv. V. 46. 1991. P. 47–106.

[3] *H. T. Croft, K. J. Falconer and R. K. Guy.* Unsolved Problems in Geometry. Springer-Verlag, New York. 1991.

[4] *G. Gallavotti and D. Ornstein.* Billiards and Bernoulli schemes. Comm. Math. Phys. V. 38. 1974. P. 83–101.

[5] *R. Guy and V. Klee.* Monthly research problems. Amer. Math. Monthly. V. 78. 1971. P. 1114.

[6] *V. Klee.* Is every polygonal region illuminable from some point?. Amer. Moth. Monthly. V. 76. 1969. P. 180.

[7] *V. Klee.* Some unsolved problems in plane geometry. Math. Magazine. V. 52. 1979. P. 131–145.

[8] *V. Klee and S. Wagon.* Old and New Unsolved Problems in Plane Geometry and Number Theory. The Math. Assoc. of America. 1991.

[9] *L. Penrose and R. Penrose.* Puzzles for Christmas. New Scientist. V. 25. 1958. P. 1580–1581, 1597.

[10] *J. Rauch.* Illuminations of bounded domains. Amer. Math. Monthly. V. 85. 1978. P. 359–361.

[11] *Ya. G. Sinai.* Dynamical systems with elastic reflections. Ergodic properties of dispersing billiards. Russ. Math. Surv. V. 25. (1970), 137–189.

[12] *S. Tabachnikov.* Billiards. Panoramas et Syntheses No. 1, 1995.

[13] *G. Tokarsky.* Polygonal rooms not illuminable from every point. Amer. Moth. Monthly. V. 102. 1995. P. 867–879.