



GEOMETRY OF REFRACTIONS AND REFLECTIONS THROUGH A BIPERIODIC MEDIUM

DOI:

[10.1137/15M1014127](https://doi.org/10.1137/15M1014127)

Document Version

Accepted author manuscript

[Link to publication record in Manchester Research Explorer](#)

Citation for published version (APA):

Glendinning, P. (2016). GEOMETRY OF REFRACTIONS AND REFLECTIONS THROUGH A BIPERIODIC MEDIUM. *SIAM Journal of Applied Mathematics*, 76(4), 1219-1238. <https://doi.org/10.1137/15M1014127>

Published in:

SIAM Journal of Applied Mathematics

Citing this paper

Please note that where the full-text provided on Manchester Research Explorer is the Author Accepted Manuscript or Proof version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version.

General rights

Copyright and moral rights for the publications made accessible in the Research Explorer are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Takedown policy

If you believe that this document breaches copyright please refer to the University of Manchester's Takedown Procedures [<http://man.ac.uk/04Y6Bo>] or contact uml.scholarlycommunications@manchester.ac.uk providing relevant details, so we can investigate your claim.



GEOMETRY OF REFRACTIONS AND REFLECTIONS THROUGH A BIPERIODIC MEDIUM*

PAUL GLENDINNING†

Abstract. The behaviour of light rays obeying Snell's Law in a medium made up of two materials with different refractive indices and which are arranged in a periodic chessboard pattern is described. The analysis is in some ways analogous to the study of rational billiards and uses a return map on one surface to prove, amongst other things, that the number of angles with which any individual ray intersects the lattice is bounded and that if the ratio of refractive indices is large enough then the dynamics can be described by interval exchange maps.

Key words. composite billiards, refraction, reflection, ergodicity, geometric ray theory, interval exchange maps

AMS subject classifications. 37E05, 78A05

1. Introduction. Geometric optics describes the passage of light through media along rays. This approximation ignores the wave-like properties of light such as interference or diffraction, but it is used to describe phenomena such as focussing and caustics. If a light ray passing through a medium with constant optical properties strikes a boundary between this medium and another with different properties so that the speed of propagation is different in each then it is either reflected or refracted. In this paper a dynamical systems approach is used to describe the geometry of rays in a medium made up of two materials arranged on an infinite chessboard. This is a first step towards understanding the paths of rays in more complex configurations which has two motivations. First, properties of the paths of light rays in complicated media are poorly understood, and there is a hope – not realized on the chessboard – of finding configurations that allow diffuse incident light to be focussed in particular directions. Second, the mathematical description of the paths has close analogies with the study of non-compact billiards and rational billiards, and it can also be seen as an extension of the interest in dynamical systems with ‘holes’. The paths of light rays modulo the symmetry of the chessboard are described by a return map which is used to prove that the number of angles with which any individual ray intersects the lattice is bounded and that if the ratio of the refractive indices is large enough then the dynamics can be described by interval exchange maps. Results for these maps can then be used to describe the ergodic properties of rays.

The mathematics of billiards describes the properties of paths formed by particles reflecting off a closed boundary so that the angle of incidence (the angle between the incoming particle path and the inward normal to the boundary at the point of contact) equals the angle of reflection (the angle between the outgoing particle path and the inward normal to the boundary), see e.g. [15]. In two dimensions the study of billiards is often described by a pair (s, θ) , or equivalently $(s, \sin \theta)$, where s is some arclength parametrization of the boundary and θ is the angle of incidence of the particle. Thus successive impacts of the particle on the boundary generate a sequence (s_n, θ_n) where

*This work was partially supported by a CNRS funded visit to the Institut Non Linéaire de Nice (INLN), France. I am grateful to Jean-Marc Gambaudo (INLN) for a series of conversations from which the idea for this research emerged and also to Stewart Johnson (Williams College) for helpful comments.

†School of Mathematics, University of Manchester, Oxford Road, Manchester M13 9PL, UK. p.a.glendinning@manchester.ac.uk.

the n^{th} collision with the boundary is simply a function of the position and angle of the previous collision. Viewed as a map this process has many nice properties depending on the geometry of the boundary on which collisions take place, the most famous being the stadium [5]. Boundaries with corners such as squares are also considered and more recently the effect of holes has been considered, where some portion of phase space (often an arc of the boundary) is removed, and any trajectory which lands in this region is deemed to be lost from the system. This is sometimes called *open billiards*. Typical questions asked in this context are: which orbits never fall in the hole, how does this change with the width of the hole, how long do orbits take to leave through a hole [7, 11] (because many billiard systems are ergodic, almost all solutions do eventually leave), and how does this change with the shape of the billiard table [3]?

More recently, and more relevant to the geometry described below, unbounded planar billiards have been analyzed, with particular emphasis on the Ehrenfest wind-tree model [6, 8, 9]. In these models polygonal obstacles are placed inside a lattice and rays move through the gaps created by a periodic continuation of this structure in both horizontal and vertical directions. Results such as the non-ergodicity of the motion can be derived in some cases, together with constraints that imply (for example) that every ray is periodic. More complicated rays, when they exist, have many similarities (from the perspective of a naive visual comparison) with sample rays described here, see Figures 1 and 6 below, and Figures 8 and 9 of [9] for example. It is clear that a more detailed comparison of the cases would be worthwhile, and I am grateful to a referee for pointing out this possible connection.

It is equally possible to think of billiards as giving the intersections of the path of a ray of light with a partially mirrored surface [15]. The law of reflection has precisely the same geometry but refraction is also possible. In this context a natural way of introducing holes is to consider the boundary as the surface separating two media with very different refractive indices. As the ratio of these indices (or equivalently the ratio of the speed of light in each medium) tends to infinity the mirror becomes perfectly reflecting, but at finite values of there is a cone of angles of incidence that pass through the surface from one medium to the other in one direction, whilst all light passes through in the other. If the two refractive indices are not equal then Snell's Law (Descartes' Law) holds, see equation (2.1) and if the ray is refracted across the surface then the ratio of the sines of the angle of incidence to the angle of refraction is constant, the ratio of the refractive indices of the two media. In one direction (from the black material with lower refractive index to the white material with higher refractive index with the convention used below) the ray always passes through, whilst in the other direction there is a critical angle θ_c such that if the angle of incidence is greater than θ_c then the ray is reflected. As the ratio of refractive indices tends to infinity almost all solutions are reflected. There has been some work on this version, which is particularly interesting as it provides the possibility of further dynamics in the second medium. In [1] the idea of a hole as being due to a non-reflective portion of the boundary is used, and the light emitted from a billiard boundary is analyzed. More interesting from the point of view of this article is the work of [4], where an annular region with perfectly reflecting outer boundary surrounding a central region with different refractive indices is considered; they call this *composite billiards*. The diffusion of light out of the central region is described and there is a discussion of how it can return after reflections on the outer boundary of the annular region.

The approach taken in this paper is similar to that of [4], except that the bounded

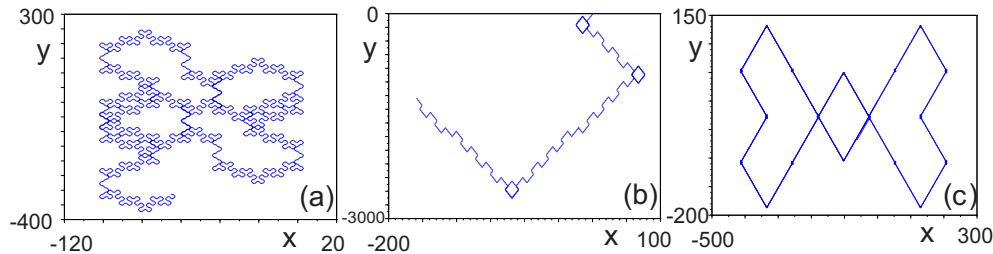


FIG. 1. Rays showing 5000 refractions and reflections with μ defined above (2.1) and $y_0 = 0$. Successive outwards intersections with the faces of ‘white’ (higher refractive index) squares are connected by straight lines; the actual refractions, when they occur, are not represented. (a) $x_0 = 0.6$; $\theta_0 = 0.1$; $\mu = 1.9$; (b) $x_0 = 0.6$; $\theta_0 = 0.01$; $\mu = 7$; and (c) $x_0 = 0.6$; $\theta_0 = 0.1$; $\mu = 7$.

property of their geometry is replaced by an infinite chessboard medium with alternating squares of two different materials. This means that even though our problem tends to billiards in a square as the ratio of the refractive indices tends to infinity, the material becomes uniform as this ratio tends to one and so the light rays are straight lines in this limit. In many ways the analogy with open billiards starts to be misleading – there is no ‘hole’ – and the problem resembles one of light in complex materials, with complicated sequences of refractions and reflections. Natural questions are what angles can light make with the boundary and how is the direction of light distributed – can it go in any direction? Is there a notion of average refractive index of the material? Note that the standard billiard in the square is not ergodic as angles are always in the set $\{\theta, \frac{\pi}{2} - \theta\}$ if reflections are on perpendicular surfaces.

The images in Figure 1 show a variety of rays illustrating the complexity (and regularity) that can be observed.

In the next section we describe the rules which determine how light rays behave in the material and derive a return map as with a billiard problem, though this factors out several of the symmetries in the problem. Rays can be divided into two separate classes (see Lemma 3.1): those which remain in one square, reflecting at every intersection with the boundary (*pure reflections*), and those that contain at least one refraction (*refracting rays*). These can have at most one reflection between sequences of refractions. We describe some properties of refracting rays in later sections. In particular we show that if the ratio of refractive indices is $\mu > 1$:

- Each ray can intersect the lines of the integer lattice at a finite number of angles.
- If $\mu > \sqrt{2}$ the dynamics of refracting rays can be described by an induced map which is an *interval exchange map*.
- If $\mu > \sqrt{2}$ then a refracting ray (modulo rotations by $\frac{m\pi}{2}$ about the midpoint of a square and reflections in vertical and horizontal lines through those centres) intersects the boundary at a dense set of points except for a countable set of incidence angles.

Note that the restriction to $\mu > \sqrt{2}$ is to make our analysis as simple as possible, and we conjecture that the last two results hold if $1 < \mu < \sqrt{2}$. The third result is analogous to the classic result for rigid rotations $x \rightarrow x + \alpha \pmod{1}$ that motion is dense on the circle except at a countable number of values of α (the rational numbers), where every point is periodic.

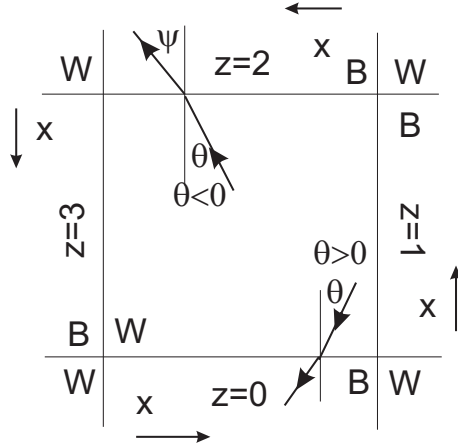


FIG. 2. Coordinates on the boundary of a white square where W indicates the neighbouring white squares and B the neighbouring black squares: each face is labelled by $z \in \{0, 1, 2, 3\}$ and the distance measured anticlockwise along each face is $x \in [0, 1)$, so the arclength from the bottom left hand corner is $z + x$.

2. Refractions and reflections on the chessboard. Suppose that the plane is tiled by unit squares with vertices on the integer lattice, labelled white (W) and black (B) alternately as on a chessboard. Suppose further that the refractive index, n_w of the material in tiles labelled by W is larger than the index n_b in the material labelled B ; think of W as water and B as air (or bromine gas). Let $\mu = n_w/n_b > 1$. If a ray of light starts at some point in W and strikes a boundary at an angle of incidence θ (measured from the normal to the surface) then Snell's Law states that the angle of refraction is ψ where

$$(2.1) \quad \mu \sin \theta = \sin \psi.$$

If $\theta > \theta_c = \sin^{-1}(1/\mu)$ there is no refracted solution and the ray is reflected back internally into W making an angle of reflection θ . Any ray from B striking the boundary with angle ψ will be refracted using (2.1), so solutions are invertible. Solutions striking the boundary at corners will be ignored here (they form a set of measure zero in the set of possible solutions).

To build up a mathematical description of the rays we will introduce coordinates on the boundary of a white square and consider a ray that strikes the boundary from inside the white square at a given point at an angle of incidence θ . Choose the white (higher refractive index material) square to be

$$(2.2) \quad S_{n,m} = \{(x, y) \mid x = n + u, y = m + v, 0 < u, v < 1\}, \quad |n| + |m| \text{ even},$$

so the black (lower refractive index material) squares have $|n| + |m|$ odd. The ray will be continued geometrically until it is next incident on the boundary of a white square from inside (i.e. the next time it strikes the boundary with a component of velocity outside the white square). We will then use the symmetries of the system to reduce the problem to a map on the set $(0, 1) \times (0, \frac{\pi}{2})$ denoting the successive impacts on the surface of the square moving from W to B and the corresponding angle of incidence modulo the symmetries.

The boundary of the white square $(0, 1)^2$ is the union of four intervals as shown in Figure 2. The arclength of the boundary measured anticlockwise from the bottom

left hand corner is $r \in [0, 4)$ and $r = z + x$ where $z \in \{0, 1, 2, 3\}$ labels the face, anticlockwise from the bottom, and $x \in [0, 1)$ is the distance along the edge measured anticlockwise as shown in Figure 2. The angle of incidence of a ray about to pass out of the square is in the interval $(-\frac{\pi}{2}, \frac{\pi}{2})$ with positive angles if the ray exiting the square has a component in the direction of decreasing x . This convention means that positive angles correspond to cases in which the scalar product of the velocity with the anticlockwise tangent is negative.

Now a ray incident on the boundary with $z \geq 1$ with arclength $r = z + x$ and angle of incidence θ is equivalent by rotational symmetry through multiples of $\frac{\pi}{2}$ about the centre of the square to a ray incident on the lower boundary at $z = 0$ with $r = x$. Moreover, if the angle of incidence is negative then by the symmetry which reflects rays about the vertical line $x = \frac{1}{2}$ we may take

$$(x, \theta) \rightarrow (1 - x, -\theta)$$

and hence restrict, modulo the symmetry, to $x \in (0, 1)$ and $\theta \in [0, \frac{\pi}{2})$.

Recall that $\mu > 1$ in (2.1). Thus a ray striking the bottom face of the white (higher refractive index) square at x with angle of incidence $\theta > 0$ will be reflected if $\theta > \theta_c$ where

$$(2.3) \quad \theta_c = \sin^{-1}(1/\mu)$$

and refracted if $\theta < \theta_c$, with angle of refraction ψ given by (2.1).

3. Reflections. Without loss of generality suppose that a ray strikes the bottom of a white square at (x, θ) with $\theta > 0$. If $\theta > \theta_c$ then the ray is reflected and will strike the boundary of the square outwards either at an angle θ again or (if it hits a side wall) $\frac{\pi}{2} - \theta$. Ignoring the measure zero boundary cases there are two possibilities:

$$\text{either } \frac{\pi}{2} - \theta > \theta_c \quad \text{or} \quad \frac{\pi}{2} - \theta < \theta_c.$$

In the former case the ray will be reflected again, and since the only angles the ray can make with the incident walls are θ and $\frac{\pi}{2} - \theta$ it will continue to be reflected for all time. This will be referred to as a *pure reflection*. By time reversal, a pure reflection cannot come from a refracted ray. Note that for pure reflections the inequalities $\theta, \frac{\pi}{2} - \theta > \theta_c$ imply that $\theta_c < \theta$ and $\frac{\pi}{2} - \theta < \frac{\pi}{2} - \theta_c$, so pure reflections are only possible if $\theta_c < \frac{\pi}{4}$, i.e. $\sin \theta_c < \frac{1}{\sqrt{2}}$ or $\mu > \sqrt{2}$.

If this ray is not a pure reflection then $\frac{\pi}{2} - \theta < \theta_c$, and hence the ray will be refracted at the first time it strikes a side wall. In this case $\theta > \frac{\pi}{4}$ and hence this always happens at the next intersection with the boundary, so every reflection that is not part of a pure reflection is followed immediately by a refraction. The effect of this reflection can be calculated by elementary geometry giving the following lemma.

LEMMA 3.1. *Suppose a ray is incident on the lower boundary of a white square at (x, θ) with $\theta > \theta_c$. If $\mu > \sqrt{2}$ and $\theta_c < \theta < \frac{\pi}{2} - \theta_c$ then the ray is a pure reflection and remains in the square for all time. Otherwise $\frac{\pi}{2} - \theta < \theta_c$ and the next intersection is with the left hand wall with coordinates (x', θ') given by*

$$(3.1) \quad x' = 1 - (x/\tan \theta), \quad \theta' = \frac{\pi}{2} - \theta,$$

after which the ray is refracted out of the square.

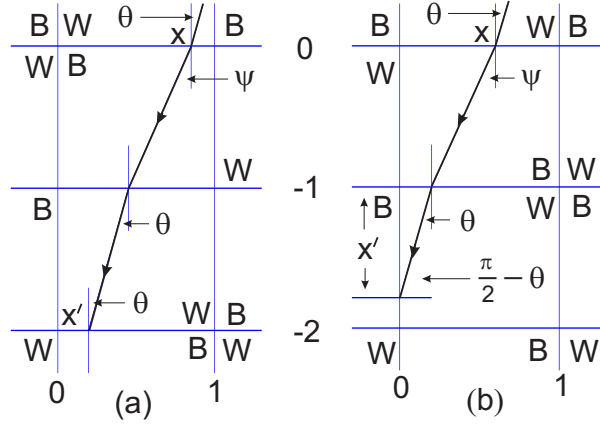


FIG. 3. Different cases of refraction showing the geometry of different rays. Each cell is labelled W (white, higher refractive index) or B (black, lower refractive index). Both rays are incident on the white squares (W) with on $y = 0$ with x -coordinate x an angle $\theta > 0$. (a) Example of ray moving down-down (DD) and striking the white square with angle θ and $x' = x - \tan \theta - \tan \psi$ measured anticlockwise along the face. (b) Example of ray moving down-left (DL) striking the $z = 3$ face with $x' = (x - \tan \psi) / \tan \theta$ and unsigned angle $\frac{\pi}{2} - \theta$. Since the line of the incident ray is in the clockwise direction at incidence the new incidence angle is negative: $\theta' = -(\frac{\pi}{2} - \theta)$.

4. Refractions. The refractions divide into four subcases as shown in Figures 3 and 4. The effect of each can be deduced directly from the geometry shown in the figures, and is summarized in the lemma below. Each individual case is described briefly in the subsequent subsections.

LEMMA 4.1. Suppose that a ray is incident on the bottom of a white square $S_{n,m}$ at (x, θ) , $0 < \theta < \theta_c$. Let ψ be the first angle of refraction

$$(4.1) \quad \psi = \sin^{-1}(\mu \sin \theta),$$

and define $\tilde{\theta}$ by

$$(4.2) \quad \tilde{\theta} = \sin^{-1}(\frac{1}{\mu} \cos \psi).$$

Then it is next incident on the boundary of a white square at (x', θ') where
(DD) if $\tan \psi + \tan \theta < x < 1$ then

$$(4.3) \quad x' = x - \tan \psi - \tan \theta, \quad \theta' = \theta,$$

and the intersection is on the face with $z = 0$ in $S_{n,m-2}$;

(DL) if $\tan \psi < x < \tan \psi + \tan \theta$ then

$$(4.4) \quad x' = (x - \tan \psi) / \tan \theta, \quad \theta' = -(\frac{\pi}{2} - \theta),$$

and the intersection is on the face with $z = 3$ in $S_{n,m-2}$;

(LD) if $\tan \psi(1 - \tan \tilde{\theta}) < x < \tan \psi$ then

$$(4.5) \quad x' = 1 - \left(1 - \frac{x}{\tan \psi}\right) / \tan \tilde{\theta}, \quad \theta' = \frac{\pi}{2} - \tilde{\theta},$$

and the intersection is on the face with $z = 0$ in $S_{n-1,m-1}$;

(LL) if $0 < x < \tan \psi(1 - \tan \tilde{\theta})$ then

$$(4.6) \quad x' = (x / \tan \psi) + \tan \tilde{\theta}, \quad \theta' = -\tilde{\theta},$$

and the intersection is on the face with $z = 3$ in $S_{n-1, m-1}$.

It is perfectly possible to calculate the domains and ranges of each of these cases explicitly, but this adds to the complexity of the description without adding to the information required to prove the main theorems for this paper, so we will not go into this level of detail here. The next four subsections sketch the derivation of the expressions given in Lemma 4.1.

4.1. Refractions: down-down. In the first case a ray incident at on the bottom face with coordinates (x, θ) , $\theta < \theta_c$, is refracted into the black region below it and then strikes the boundary of the white region directly below it, is refracted through this and is then incident on the bottom of this white square with coordinates (x', θ') . Thus it goes down through the black square and then down again through the white, and we label this *down-down*, or *DD*. After the first refraction, the angle of refraction (in the black square) is ψ defined in (4.1) and provided $\tan \psi < x$ then the intersection with the bottom of this black square is at $x - \tan \psi$. The refraction into the white square below returns the angle to θ and this will then be incident on the bottom face of this white square if $x > \tan \psi + \tan \theta$ with (x', θ') given by (4.3).

4.2. Refractions: down-left. In Figure 3b the refracted ray passes down through the black square but after entering the white square below it, the ray is incident on the left face of this square. This occurs if $\tan \psi < x < \tan \psi + \tan \theta$. In this case the angle of incidence is $-(\frac{\pi}{2} - \theta)$, with the minus sign because the incident ray is ‘behind’ the direction of increasing x on the left boundary, and an elementary geometric calculation shows that the new coordinate on the left hand boundary is (x', θ') where the arclength is $r' = 3 + x'$ with (x', θ') given by (4.4).

4.3. Refractions: left-down. If $x < \tan \psi$ then the refracted ray strikes the left hand boundary of the black square at an angle $\frac{\pi}{2} - \psi$ and is refracted through into the neighbouring white square making an angle $\tilde{\theta}$ with the normal to the boundary in the white square, where $\mu \sin \tilde{\theta} = \cos \psi$, i.e. $\tilde{\theta}$ is defined by (4.2).

It strikes the boundary from black to white a distance $y = x / \tan \psi$ below the upper left hand corner of the black square and so if $(x / \tan \psi) + \tan \tilde{\theta} > 1$ it is incident on the bottom of the white square at x' where $\tan \tilde{\theta} = (1 - y) / (1 - x')$ at an angle $\frac{\pi}{2} - \tilde{\theta}$, so (x', θ') are given by (4.5).

4.4. Refractions: left-left. If $x < \tan \psi$ and $(x / \tan \psi) + \tan \tilde{\theta} < 1$ then the ray is refracted through the white square to the left of the black square and is incident on the left face with angle of incidence $-\tilde{\theta}$ as shown in Figure 4b with (x', θ') given by (4.5).

5. Finite θ orbits. Given an angle of incidence θ in a white square with angle of refraction ψ , the ray may strike the side of a black square with angle $\frac{\pi}{2} - \psi$ and then be refracted into a white square with angle $\tilde{\theta}$. Thus we have associated an angle of incidence θ with an angle of incidence $\tilde{\theta}$. But on entry in the white square with angle $\tilde{\theta}$ the ray may be incident on a different side of the white square with angle of incidence $\frac{\pi}{2} - \tilde{\theta}$. Can this generate new angles of incidence after further refractions?

It will turn out that for all $\mu > 1$ rays have a finite number of possible angles that can be generated in this way, and hence that the θ behaviour always involves just a

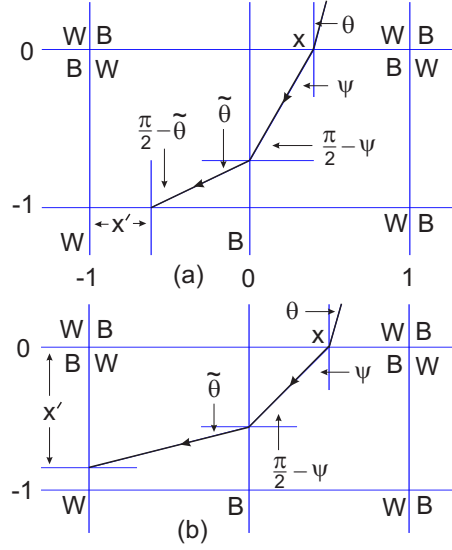


FIG. 4. As Figure 3. (a) Example of ray moving left-down (LD) and striking the white square with angle $\frac{\pi}{2} - \tilde{\theta}$ and $x' = 1 - (1 - x/\tan\psi)/\tan\theta$ measured anticlockwise along the face. (b) Example of ray moving left-left (LL) striking the $z = 3$ face with $x' = (x/\tan\psi) + \tan\theta$ and unsigned angle $\tilde{\theta}$. Since the line of the incident ray is in the clockwise direction at incidence the new incidence angle is negative: $\theta' = -\tilde{\theta}$.

finite number of θ values. This will make it possible to reduce the two-dimensional map of the previous section to a one-dimensional piecewise linear map.

THEOREM 5.1. *If $\mu > \mu_n$,*

$$(5.1) \quad \mu_n = \sqrt{\frac{n+1}{n}}$$

$n = 1, 2, \dots$, then given any initial value of θ the ray through (x, θ) takes at most $2n + 2$ different values of the angle variable.

Proof: Given an initial $\theta_0 < \theta_c$ then by Lemma 4.1 the image angle, i.e. the next angle of incidence from inside a white square modulo the symmetry, has $\theta'_0 \in \{\theta_0, \frac{\pi}{2} - \theta_0, \tilde{\theta}_0, \frac{\pi}{2} - \tilde{\theta}_0\}$ and each of these in turn can refract if they are less than θ_c . Recall that by definition

$$(5.2) \quad \mu \sin \theta_0 = \sin \psi_0, \quad \mu \sin \tilde{\theta}_0 = \cos \psi_0 = \sin(\frac{\pi}{2} - \psi_0),$$

and $\theta_0 < \theta_c$ by assumption, $\tilde{\theta}_0 < \theta_c$ as $\sin \tilde{\theta}_0 = \cos \psi_0 / \mu < 1/\mu$, and hence $\frac{\pi}{2} - \theta_0$ and $\frac{\pi}{2} - \tilde{\theta}_0$ are both greater than $\frac{\pi}{2} - \theta_c$.

Consider $\theta'_0 = \frac{\pi}{2} - \tilde{\theta}_0 > \frac{\pi}{2} - \theta_c$. If $\theta'_0 > \theta_c$ then the ray will be reflected, and the next outward intersection with the boundary of the square has angle $\tilde{\theta}_0$ again, so there is no new refracting angle. Note that $\theta'_0 > \theta_c$ if $\frac{\pi}{2} - \theta_c > \theta_c$, i.e. if $\theta_c < \frac{\pi}{4}$ or $\mu > \sqrt{2} = \mu_1$.

Now suppose that $\theta'_0 < \theta_c$. This can create a new angle of incidence if there exists ψ_1 and $\tilde{\theta}_1$ such that

$$(5.3) \quad \mu \sin \theta_1 = \sin \psi_1, \quad \mu \sin \tilde{\theta}_1 = \cos \psi_1.$$

Thus the new angle $\tilde{\theta}_1$ exists provided $(\cos \psi_1)/\mu < 1$ which is automatically satisfied if $\mu \sin \theta_1 < 1$.

It will actually be easier to consider the new refracted angle ψ_1 rather than $\tilde{\theta}_1$. Now (5.3) implies

$$\begin{aligned} \sin \psi_1 &= \mu \sin \theta_1 = \mu \sin(\frac{\pi}{2} - \tilde{\theta}_0) = \mu \cos \tilde{\theta}_0 \\ (5.4) \quad &= \sqrt{\mu^2 - \mu^2 \sin^2 \tilde{\theta}_0} = \sqrt{\mu^2 - \cos^2 \psi_0} \\ &= \sqrt{\mu^2 - 1 + \sin^2 \psi_0} \end{aligned}$$

and so defining

$$(5.5) \quad u_0 = \sin \psi_0, \quad u_1 = \sin \psi_1$$

this implies that

$$(5.6) \quad u_1 = \sqrt{\mu^2 - 1 + u_0^2}$$

provided $u_1 < 1$.

Now consider the possibility of generating a new angle of incidence via the angle $\tilde{\theta}_{-1} = \frac{\pi}{2} - \theta_0$ which is also greater than $\frac{\pi}{2} - \theta_c$. The equivalent of (5.2) for the new angle θ_{-1} uses the angle ψ_{-1} defined by

$$(5.7) \quad \mu \sin \theta_{-1} = \sin \psi_{-1}, \quad \mu \sin \tilde{\theta}_{-1} = \cos \psi_{-1}.$$

Thus

$$\begin{aligned} \sin \tilde{\psi}_{-1} &= \sqrt{1 - \cos^2 \psi_{-1}} = \sqrt{1 - \mu^2 \sin^2 \tilde{\theta}_{-1}} = \sqrt{1 - \mu^2 \sin^2(\frac{\pi}{2} - \theta_0)} \\ (5.8) \quad &= \sqrt{1 - \mu^2 \cos^2 \theta_0} = \sqrt{1 - \mu^2 + \mu^2 \sin^2 \theta_0} \\ &= \sqrt{1 - \mu^2 + \sin^2 \psi_0}. \end{aligned}$$

Thus setting

$$(5.9) \quad u_{-1} = \sin \psi_{-1}$$

with u_0 as in (5.5),

$$(5.10) \quad u_{-1} = \sqrt{1 - \mu^2 + u_0^2}$$

provided $\tilde{u}_{-1} < 1$.

On the face of it this suggests that for each pair $(\theta_1, \tilde{\theta}_1)$, the pair $(\frac{\pi}{2} - \theta_1, \frac{\pi}{2} - \tilde{\theta}_1)$ can exist and a further two angles can be defined via (5.6) and (5.10). However, G is defined by

$$(5.11) \quad G(u) = \sqrt{\mu^2 - 1 + u^2}$$

so that (5.6) is $u_1 = G(u_0)$, then

$$(5.12) \quad G^{-1}(u) = \sqrt{1 - \mu^2 + u^2}$$

and so (5.10) is $u_{-1} = G^{-1}(u_0)$. Thus given u_0 we can generate new angles by considering the forwards and backwards iterates of G , i.e.

$$(5.13) \quad \mathcal{G} = \{u_k \in [0, 1] \mid u_k = G^{k-1}(u_0), \ k \in \mathbb{Z}\}.$$

Each element of \mathcal{G} generates a pair of new angles, i.e. u_0 generates $(\theta_0, \tilde{\theta}_0)$, u_1 (if it exists) generates $(\frac{\pi}{2} - \theta_0, \theta_1)$, u_{-1} generates $(\frac{\pi}{2} - \tilde{\theta}_0, \theta_{-1})$ and so on. Thus if the cardinality of \mathcal{G} is $g < \infty$ then the number of possible angles of a ray with incidence angle θ_1 is

$$(5.14) \quad 2 + 2g$$

as the first determines two angles, whilst the subsequent iterations adds a couple more until the last possible case, where the corresponding $\frac{\pi}{2} - \theta_n$ and $\frac{\pi}{2} - \theta_{-n}$ need to be included.

If $\mu > 1$ then G is monotonic increasing and $G(u) > u$ so it has no fixed points and (u_k) is a strictly increasing sequence. Moreover by direct calculation

$$(5.15) \quad G^n(0) = \sqrt{n(\mu^2 - 1)}$$

and so $G^n(0) \leq 1$ provided $\mu^2 \leq (n+1)/n$. Let $\mu_n = \sqrt{\frac{n+1}{n}}$ and note $\mu_1 = \sqrt{2}$. If $\mu > \mu_1$ then $g = 1$ and there are four angles possible. If $\mu_n < \mu < \mu_{n-1}$, $g \leq n$ and so the maximum number of angles that can be generated is $2n + 2$. As $\mu \rightarrow 1$ from above more and more possible angles become possible, converging onto solutions with more and more angles close to zero – in other words the limiting solution rays are asymptotically straight (horizontal or vertical). \square

6. A factorized return map. Given a ray which is not a pure reflection incident on the face of a white (higher refractive index) square from inside the square at (x, θ) , Lemmas 3.1 and 4.1 show how to determine the new coordinates (x', θ') of the refracted or reflected ray at the next incidence on a face from within a white square. The lemmas hold only for positive angles of incidence. The equivalent results for negative angles can be obtained using symmetry: let $S(x, \theta) = (1 - x, -\theta)$ then if $(x, \theta) \in [0, 1] \times [0, \frac{\pi}{2}]$ let $U(x, \theta)$ denote the pair (x', θ') of Lemmas 3.1 and 4.1. Then the full return map from faces to faces of the square can be defined using symmetry:

$$(6.1) \quad T(x, \theta) = \begin{cases} U(x, \theta) & \text{if } \theta \geq 0 \\ SUS(x, \theta) & \text{if } \theta < 0. \end{cases}$$

To simplify further we can use the symmetry S to construct a map from positive angles to positive angles: if $U(x, \theta) = (x', \theta')$ for $\theta > 0$ then define

$$(6.2) \quad F(x, \theta) = \begin{cases} U(x, \theta) & \text{if } \theta' \geq 0 \\ SU(x, \theta) & \text{if } \theta' < 0. \end{cases}$$

Then $F : [0, 1] \times [0, \frac{\pi}{2}] \rightarrow [0, 1] \times [0, \frac{\pi}{2}]$ (and we are ignoring rays that go through corners) and a tiresome exercise in induction shows that $F^k(x, \theta) \in \{T^k(x, \theta), ST^k(x, \theta)\}$ for rays avoiding corners. In other words the map (6.2) encapsulates the dynamics of successive outgoing intersections with the boundary of the square up to the symmetry S .

The reflections (3.1) already map positive angles to positive angles and the new map (6.2) changes the rules of two of the four refractions of Lemma 4.1 which become, using the abbreviation DD for ‘down-down’, DL for ‘down-left’ etc and adding RR

for the reflection of Lemma 3.1:

$$(6.3) \quad \begin{array}{ll} DD & x' = x - \tan \psi - \tan \theta \quad \theta' = \theta \\ DL & x' = 1 - (x - \tan \psi) / \tan \theta \quad \theta' = \frac{\pi}{2} - \theta \\ LD & x' = 1 - (1 - (x / \tan \psi)) / \tan \tilde{\theta} \quad \theta' = \frac{\pi}{2} - \tilde{\theta} \\ LL & x' = 1 - \tan \tilde{\theta} - (x / \tan \psi) \quad \theta' = \tilde{\theta} \\ RR & x' = 1 - (x / \tan \theta) \quad \theta' = \frac{\pi}{2} - \theta \end{array}$$

with ψ and $\tilde{\theta}$ defined in (4.1) and (4.2). These five maps have their domains in different parts of the phase space. For DD the map is defined on

$$(6.4) \quad DD = \{(x, \theta) \mid \tan \psi + \tan \theta < x < 1, 0 < \theta < \hat{\theta}\}$$

where if $\hat{\psi}$ is defined in terms of $\hat{\theta}$ via (4.1), i.e. $\hat{\psi}$ is the refracted ray in B if the angle of incidence from W is $\hat{\theta}$, then

$$(6.5) \quad \tan \hat{\psi} + \tan \hat{\theta} = 1.$$

For DL

$$(6.6) \quad DL = \{(x, \theta) \mid \tan \psi < x < \min(1, \tan \psi + \tan \theta), 0 < \theta < \theta_{cc}\},$$

where θ_{cc} is the value at which the corresponding refracted angle, ψ_{cc} , satisfies $\tan \psi_{cc} = 1$, i.e.

$$(6.7) \quad \theta_{cc} = \sin^{-1}(1/(\mu\sqrt{2})), \quad \psi_{cc} = \frac{\pi}{4}, \quad \tilde{\theta}_{cc} = \theta_{cc}.$$

The latter follows because if the refracted angle is $\frac{\pi}{4}$ then this ray hits a side of the square it does so with angle $\frac{\pi}{2} - \frac{\pi}{4} = \frac{\pi}{4}$ and hence $\tilde{\theta}$ is equal to θ in this case. At $x = 1$, the values of θ for DL lie between $\hat{\theta}$ and θ_{cc} and so if $\theta \in (\hat{\theta}, \theta_{cc})$ the minimum in the definition of DL is 1, and below this range it is the other expression.

For LD

$$(6.8) \quad LD = \{(x, \theta) \mid (1 - \tan \tilde{\theta}) \tan \psi < x < \min(1, \tan \psi), \theta'_o < \theta < \theta_o\}$$

where $\theta'_o = 0$ if $\mu \geq \sqrt{2}$ and $\mu \sin \theta'_o = \sqrt{1 - \frac{\mu^2}{2}}$ if $1 < \mu < \sqrt{2}$, i.e. $\tan \tilde{\theta}'_o = 1$; θ_o is defined by

$$(6.9) \quad (1 - \tan \tilde{\theta}_o) \tan \psi_o = 1;$$

and ψ_o and $\tilde{\theta}_o$ are defined from θ_o by (4.1) and (4.2). The minimum in (6.8) equals 1 in (θ_{cc}, θ_o) and $\tan \psi$ below this interval.

For LL the domain is

$$(6.10) \quad LL = \{(x, \theta) \mid 0 < x < \min(1, (1 - \tan \tilde{\theta}) \tan \psi), 0 < \theta < \theta_c\}$$

and splits up into two regions naturally: if $0 < \theta < \theta_o$ it is a triangular region with $x < (1 - \tan \tilde{\theta}) \tan \psi$ whilst if $\theta_o < \theta < \theta_c$ it is rectangular with $0 < x < 1$.

Finally the reflected domain RR containing only those intersections that are not pure reflections is defined by

$$(6.11) \quad RR = \{(x, \theta) \mid 0 < x < 1, \max(\theta_c, \frac{\pi}{2} - \theta_c) < \theta < \frac{\pi}{2}\}.$$

The image of each of these regions is not hard to calculate, but the details are not needed in the analysis below so we will not go further into this description.

Note that the return maps of (6.2) have been obtained by factorizing out the symmetries of the chessboard generated by rotations through $\frac{\pi}{2}$ and reflections in half integer horizontal and vertical lines and so the interpretation of solutions in terms of directions travelled and their ‘true’ positions is not transparently obvious from solutions of the factorized map, on the other hand the map is considerably easier to analyze than the full map.

7. Induced maps for $\mu > \sqrt{2}$. If $\mu > \sqrt{2}$ then by Theorem 5.1 each orbit generates at most four incidence angles, and so the induced map (6.3) for an initial condition (\cdot, θ) is described by a one-dimensional map with discontinuities obtained by combining the maps for the four different possible angles together.

In this case the return map is (6.3) if those angles that give the classic billiards of pure reflections are ignored. Moreover, without loss of generality the initial angle θ may be chosen so that the four possible angles are

$$(7.1) \quad \theta < \tilde{\theta} \leq \theta_c < \frac{\pi}{2} - \tilde{\theta} \leq \frac{\pi}{2} - \theta.$$

and we will only consider the general case where inequality does not hold here. To convert the entire map to a map of the interval identify the interval $(0, 1)$ with $\{(x, \phi) \mid \phi = \theta\}$, the interval $(1, 2)$ with $\{(x, \phi) \mid \phi = \tilde{\theta}\}$, the interval $(2, 3)$ with $\{(x, \phi) \mid \phi = \frac{\pi}{2} - \tilde{\theta}\}$ and $(3, 4)$ with $\{(x, \phi) \mid \phi = \frac{\pi}{2} - \theta\}$. Then (6.3) induces a one dimensional map D with discontinuities on the interval $[0, 4]$ where (choosing the easier induced maps first):

$$(7.2) \quad \begin{aligned} D(x) &= 1 - (x - 3)/\tan(\frac{\pi}{2} - \theta) \\ &= 1 - (x - 3)\tan\theta \quad \text{if } x \in [3, 4] \end{aligned}$$

since the initial angle is $\frac{\pi}{2} - \theta$ by the identification above, and the image has angle $\frac{\pi}{2} - (\frac{\pi}{2} - \theta) = \theta$. Similarly

$$(7.3) \quad D(x) = 2 - (x - 2)\tan\tilde{\theta} \quad \text{if } x \in [2, 3]$$

On $[0, 1]$ and $[1, 2]$ the situation is more complicated as, at first glance, any one of four maps might be applied, depending on the values of θ and $\tilde{\theta}$ and how they intersect the different regions DD , DL , LD and LL , and the appropriate shifts need to be made to ensure that the images lie in the correct intervals. Thus,

$$(7.4) \quad D(x) = \begin{cases} x - \tan\psi - \tan\theta & \text{if } x \in (\tan\psi + \tan\theta, 1) & (DD) \\ 4 - (x - \tan\psi)\cot\theta & \text{if } x \in (\tan\psi, \tan\psi + \tan\theta) & (DL) \\ 3 - (1 - x\cot\psi)\cot\tilde{\theta} & \text{if } x \in ((1 - \tan\tilde{\theta})\tan\psi, \tan\psi) & (LD) \\ 2 - \tan\tilde{\theta} - x\cot\psi & \text{if } x \in (0, (1 - \tan\tilde{\theta})\tan\psi) & (LL) \end{cases}$$

where the shifts have been chosen to ensure that the images lie in the intervals with the appropriate angle variable consistent with (6.3). In the case $\sin\theta > \frac{1}{\mu\sqrt{2}}$ the region DD is empty and the right hand end-point of the range of DL is $x = 1$. This latter case occurs if $\tan\theta + \tan\psi > 1$, i.e. if $\theta > \hat{\theta}$ (cf. previous sections).

If $x \in [1, 2]$, the situation is similar, with the roles of θ and $\tilde{\theta}$ reversed and the refracted angle corresponding to $\tilde{\theta}$ is $\frac{\pi}{2} - \psi$. However, $\tan\psi < 1$ as $\mu > \sqrt{2}$,

$\tan \tilde{\psi} = \cot \psi > 1$ and hence the conditions for DD and DL are never satisfied. Moreover, the equivalent of the map LL in $(1, 2)$ is

$$1 - \tan \theta - (x - 1) \tan \psi$$

and the boundary of LL and LD is the point at which this takes the value zero (if this exists). But if $1 - \tan \theta - \tan \psi > 0$, i.e. $\theta < \hat{\theta}$ defined in (6.5) there is no zero in $(1, 2)$ and the only branch of the map that can be applied is LL , so

$$(7.5) \quad D(x) = 1 - \tan \theta - (x - 1) \tan \psi \quad \text{if } x \in (1, 2). \quad (LL)$$

On the other hand, if $\theta > \hat{\theta}$ then there is a zero at x' where $(x' - 1) \tan \psi = 1 - \tan \theta$ and then

$$(7.6) \quad D(x) = \begin{cases} 1 - \tan \theta - (x - 1) \tan \psi & \text{if } x \in (1, x') \quad (LL) \\ 4 - (1 - (x - 1) \tan \psi) \cot \theta & \text{if } x \in (x', 2). \quad (LD) \end{cases}$$

This return map can, fortunately, be simplified further, but, unfortunately, there are two cases that need to be considered separately.

7.1. Case A. If $0 < \theta < \hat{\theta}$ where $\tan \hat{\theta} + \tan \hat{\psi} = 1$, see (6.5), the return map has seven branches defined by (7.2), (7.3), all four branches of (7.4) and (7.5). We will refer to this as case A.

For convenience define

$$x_0 = 0, \quad x_1 = (1 - \tan \tilde{\theta}) \tan \psi, \quad x_2 = \tan \psi, \quad x_3 = \tan \psi + \tan \theta, \quad x_4 = 1,$$

and

$$I_n = (x_{n-1}, x_n), \quad n = 1, \dots, 4.$$

If the interval $[0, 1)$ is divided into subintervals I_k then an interval exchange map [18] is a bijection that permutes the intervals. These have some beautiful properties (see below) and although the map D defined above is not an interval exchange map, it is possible to use iterates of D to create interval exchange maps. Interval exchange maps have been studied in many contexts since the mid-1970s [2, 10, 16] and this will make the analysis of the induced map H relatively simple, though there remain some problems in interpreting results for H in terms of the dynamics of the original map on surfaces. The connection with interval exchange maps suggests that there are much deeper ways of understanding the geometry of light rays through complicated media involving Teichmüller spaces and foliations on surfaces [13, 17] that we return to in the conclusion. Recall the definition of $\hat{\theta}$ from (6.5).

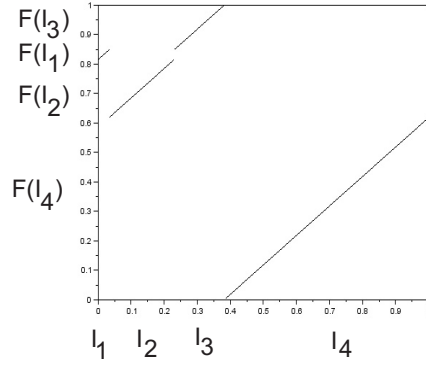
PROPOSITION 7.1. *Suppose that $0 < \theta < \hat{\theta}$ and let (I_n) , $n = 1, \dots, 4$, and D be as above. Then the induced map*

$$(7.7) \quad H(x) = \begin{cases} D^2(x) & \text{if } x \in I_1 \\ D^3(x) & \text{if } x \in I_2 \\ D^2(x) & \text{if } x \in I_3 \\ D(x) & \text{if } x \in I_4 \end{cases}$$

extended to $[0, 1)$ using continuity from the right is an interval exchange map.

Proof: Consider the map on each of the intervals I_k as shown in Fig 5. On I_4 , the first equation of (7.4) holds and

$$(7.8) \quad D(I_4) = (0, 1 - \tan \psi - \tan \theta).$$

FIG. 5. Induced map H with $\mu = 1.5$ and $\theta = 0.15$.

Now note that using (7.2), $D(3, 4) = (1 - \tan \theta, 1)$, so if $x \in I_3$, i.e. $\tan \psi < x < \tan \psi + \tan \theta$ then the map labelled DL in (7.4) is applied and the left hand limit ($x = \tan \psi$) is 4 and the right hand limit is 3, hence the image of this region is the whole of $(3, 4)$ and so with a little more calculation the second iterate is

$$(7.9) \quad D^2(x) = x + 1 - \tan \psi - \tan \theta, \quad x \in (\tan \psi, \tan \psi + \tan \theta).$$

and at the left end-point D^2 takes the value $1 - \tan \theta$ and at the right end-point it takes the value 1. Thus

$$(7.10) \quad D^2(I_3) = (1 - \tan \theta, 1).$$

Now consider $x \in I_2$, with D defined by the third equation of (7.4) labelled LD . Thus $D(I_2) = (2, 3)$ and D has slope $\cot \psi \cot \tilde{\theta}$. $D^2(I_2) = D(2, 3)$ where D is defined by (7.3) and has slope $-\tan \tilde{\theta}$, so D^2 has slope $-\cot \psi$ and $D^2(I_2) \subseteq (1, 2)$, indeed, a quick calculations shows $D^2(I_2) = (2 - \tan \tilde{\theta}, 2)$. In the interval $(1, 2)$ D is defined by the map (7.5) (with slope $\tan \psi$) and so the image lies in $(0, 1)$ and D^3 has slope 1. Moreover,

$$(7.11) \quad D^3(I_2) = (1 - \tan \theta - \tan \psi, 1 - \tan \theta - \tan \psi + \tan \tilde{\theta} \tan \psi)$$

and, putting three small calculations together, if $x \in I_2$ then

$$(7.12) \quad D^3(x) = x + 1 - 2 \tan \psi - \tan \theta + \tan \tilde{\theta} \tan \psi.$$

Finally consider $x \in I_1$ which is the region labelled LL in (7.4), so the image of the left end-point ($x = 0$) is $2 - \tan \tilde{\theta}$ and the image of the right end-point is 1, whilst the slope of the map is $\cot \psi$. The image is therefore in $(1, 2)$ where (7.5) applies and so the second iterate of the right hand end point of LL in $(0, 1)$ is the image of $x = 1$ under (7.5) which is $1 - \tan \theta$, whilst the second iterate of the left hand end point is $1 - \tan \theta - \tan \psi + \tan \tilde{\theta} \tan \psi$ as in the previous paragraph. The slope of D^2 is the product of $\cot \psi$ (from LL in (7.4)) and $\tan \psi$ from (7.5), i.e. it is 1 again and

$$(7.13) \quad D^2(I_1) = (1 - \tan \theta - \tan \psi + \tan \tilde{\theta} \tan \psi, 1 - \tan \theta)$$

and if $x \in I_1$ then

$$(7.14) \quad D^2(x) = x + 1 - \tan \theta - (1 - \tan \tilde{\theta}) \tan \psi.$$

Thus on each interval I_k $H(x)$ is a translation, and since the images do not intersect H is an interval exchange map. \square

We will return to the details of the dynamics of H in section 8.

7.2. Case B. If $\hat{\theta} < \theta < \theta_{cc}$, where $\hat{\theta}$ and θ_{cc} are defined in (6.5) and (6.7), the return map has seven branches defined by (7.2), (7.3), all four branches of (7.4) and (7.5). We will refer to this as case B. In this case the return map is defined by (7.2), (7.3), the latter three branches of (7.4) and the two branches of (7.6), separated by x' where $(x' - 1) \tan \psi = 1 - \tan \theta$. Define

$$(7.15) \quad y_0 = 0, \quad y_1 = (1 - \tan \tilde{\theta}) \tan \psi, \quad y_2 = Y, \quad y_3 = \tan \psi, \quad y_4 = 1.$$

We aim to choose Y so that an appropriately defined induced map is an interval exchange map on the four intervals $J_n = [y_n - 1, y_n)$, $n = 1, \dots, 4$.

PROPOSITION 7.2. *Suppose that $\hat{\theta} < \theta < \theta_{cc}$ and let (J_n) , $n = 1, \dots, 4$, and D be as above. If*

$$(7.16) \quad Y = \tan \psi (1 - \tan \tilde{\theta}) + \tan \theta + \tan \psi - 1$$

then the induced map

$$(7.17) \quad K(x) = \begin{cases} D^2(x) & \text{if } x \in J_1 \\ D^4(x) & \text{if } x \in J_2 \\ D^3(x) & \text{if } x \in J_3 \\ D^2(x) & \text{if } x \in J_4 \end{cases}$$

extended to $[0, 1)$ using continuity from the right is an interval exchange map.

Proof: First note that $D(0) = 2 - \tan \tilde{\theta}$ and we claim that $D(0) < x'$, or equivalently $(D(0) - 1) \tan \psi < (x' - 1) \tan \psi$. This is equivalent to $\tan \psi < \frac{1 - \tan \tilde{\theta}}{1 - \tan \theta}$ (since $\tilde{\theta} < \frac{\pi}{4}$). But by definition $\theta < \tilde{\theta} < \frac{\pi}{4}$ the ratio on the right hand side is clearly greater than one and hence this inequality is automatically satisfied and the condition $D(0) < x'$ is verified.

This condition implies that $D(J_1) = (1, 2 - \tan \tilde{\theta}) \subset (1, x')$. As before, $D(2, 3) = (2 - \tan \tilde{\theta}, 2)$ and so this contains x' , and by (7.3) the preimage of x' , $y' \in (2, 3)$, satisfies

$$y' - 2 = (2 - x') \cot \tilde{\theta}$$

and again as before, $D(J_2) = (2, 3)$ and the preimage of y' is Y – the point to be defined in (7.15) where

$$3 - (1 - Y \cot \psi) \cot \tilde{\theta} = 2 + (2 - x') \cot \tilde{\theta}$$

or

$$\tan \tilde{\theta} - (1 - Y \cot \psi) = 1 - (x' - 1).$$

Multiplying through by $\tan \psi$ and substitute for $(x' - 1)$:

$$\tan \psi \tan \tilde{\theta} + Y = 2 \tan \psi - (1 - \tan \theta)$$

so Y is as defined in the statement of the proposition, (7.16).

With these preliminaries the calculations are almost identical to those in Case A, except the induced map is (7.17): in all cases the slope of the induced map is one ($-\cot \psi \times -\tan \psi$ in J_1 , $\cot \psi \tan \tilde{\theta} \times -\tan \tilde{\theta} \times \tan \psi \cot \theta \times -\tan \theta$ in J_2 , $\cot \psi \tan \tilde{\theta} \times -\tan \tilde{\theta} \times -\tan \psi$ in J_3 and $-\cot \theta \times -\tan \theta$ in J_4 , and images do not overlap, hence K is an interval exchange map. With a bit more work that we leave to the reader the linages of J_1 to J_4 are rearranged into the order

$$(7.18) \quad K(J_3) \quad K(J_1) \quad K(J_4) \quad K(J_2)$$

which, as we shall see shortly, is enough together with the lengths of the intervals J_n to specify the map completely. \square

8. The induced map H and interval exchange maps: case A. As described in section 7.1, an interval exchange map does exactly what its name suggests. The interval $[0, 1)$ is partitioned into $N > 1$ subintervals $I_n = [x_{n-1}, x_n)$, $n = 1, \dots, N$, with $x_0 = 0$ and $x_N = 1$. An interval exchange map is a bijection of $[0, 1)$ which acts on each interval I_n as a translation $x + \alpha_n$, for some translation vector $\alpha_n \in \mathbb{R}^N$. This means that the map simply permutes the order of the intervals. In the case of H (7.7), the intervals I_1 to I_4 are in ascending order (I_1 is to the left of I_2 etc.) whilst their images lie in the order $H(I_4)$, $H(I_2)$, $H(I_1)$, $H(I_3)$, so the *monodromy invariant* of H is the permutation

$$(8.1) \quad \pi_M = (\begin{smallmatrix} 4 & 2 & 1 & 3 \end{smallmatrix})$$

describing the permutation of the order of the original partition to its image. A monodromy invariant π is *reducible* if there exists $k < N$ such that $\pi(r) \leq k$ for all $r \leq k$, otherwise it is *irreducible*. Clearly (8.1) is irreducible. Rather than use the translation vector, it is standard to define the interval exchange map in terms of the length vector $\lambda \in \mathbb{R}^N$, where $\lambda_n = x_n - x_{n-1}$ [18]. In the case of H , where $N = 4$,

$$(8.2) \quad \lambda = ((1 - \tan \tilde{\theta}) \tan \psi, \tan \tilde{\theta} \tan \psi, \tan \theta, 1 - \tan \theta - \tan \psi).$$

A length vector is *rationally dependent* if there exists $b \in \mathbb{Q}^N$, $b \neq 0$ such that

$$(8.3) \quad \sum_1^N b_k \lambda_k = 0.$$

Note that since the sum of the λ_k equals to one, one of the components, λ_r say, can be eliminated and so (8.3) is equivalent to the existence of $r \in \{1, \dots, N\}$ and $c \in \mathbb{Q}^{N-1}$, $c \neq 0$, such that

$$(8.4) \quad \sum_{k \neq r} c_k \lambda_k = 1.$$

Recall that by definition, $0 < \tan^{-1} \theta < \frac{1}{\mu\sqrt{2}}$ and $\mu^2 > 2$, so as θ varies we may expect there to be many values at which λ is rationally dependent, though typical values of θ will be rationally independent. Finally we say that an interval exchange map f satisfies the *Keane condition* if for all $n \in \{0, \dots, N-1\}$ and $r \geq 1$

$$(8.5) \quad f^r(x_i) \notin \{x_1, \dots, x_{N-1}\}.$$

Note that if $p = \pi_M^{-1}(1)$ then $f(x_{p-1}) = x_0 = 0$ (in our case, $H(x_3) = 0$), which explains why x_0 is not included in the right hand side of (8.5).

A map $f : I \rightarrow I$ is *minimal* if every orbit is dense in I . If I is an interval this implies that f has no periodic orbits. The following is one of the major results for interval exchange maps, see [18] for example.

THEOREM 8.1. *Suppose $f : [0, 1) \rightarrow [0, 1)$ is an interval exchange map with monodromy invariant π_M and length vector λ .*

(a) *If π_M is irreducible and λ is not rationally dependent then f satisfies the Keane condition (8.5).*

(b) *f is minimal if and only if f satisfies the Keane condition.*

This also implies that almost all interval exchange maps are minimal [18], although of course our maps H are not generic so this is harder to apply directly. Indeed, from (8.2) and (8.4) λ is rationally dependent if there exist rational q_1, q_2 and q_3 at least one of which is non-zero such that

$$(8.6) \quad q_1(1 - \tan \tilde{\theta}) \tan \psi + q_2 \tan \tilde{\theta} \tan \psi + q_3 \tan \theta = 1$$

or equivalently rationals c_1, c_2 and c_3 at least one of which is non-zero such that $c_1 \tan \psi + c_2 \tan \tilde{\theta} \tan \psi + c_3 \tan \theta = 1$. Using the parametrization $\alpha = \tan \psi$ this becomes

$$(8.7) \quad c_1 \alpha + \frac{c_2 \alpha}{\sqrt{\mu^2 - 1 + \mu^2 \alpha^2}} + \frac{c_3 \alpha}{\sqrt{(\mu^2 - 1) \alpha^2 + \mu^2}} = 1.$$

This will enable us to prove the following lemma.

LEMMA 8.2. *If $\mu > \sqrt{2}$ then for all but a countable set of $\theta \in (0, \hat{\theta})$ all orbits of (7.7) are minimal.*

Proof: By Theorem 8.1 we need only show that the values of α which satisfy (8.7) for rational c_k not all identically zero is countable.

Fix $\mu > \sqrt{2}$ and let $p_1(\alpha) = \mu^2 - 1 + \mu^2 \alpha^2$ and $p_2(\alpha) = (\mu^2 - 1) \alpha^2 + \mu^2$. Then by basic manipulation, if α satisfies (8.7) then it also satisfies

$$(8.8) \quad -4c_2^2 c_3^2 \alpha^4 p_1(\alpha) p_2(\alpha) + [(1 - c_1 \alpha)^2 p_1 p_2 - c_2 \alpha^2 p_2 - c_3 \alpha^2 p_1]^2 = 0.$$

For each of the countable set of rationals $(c_1, c_2, c_3) \neq (0, 0, 0)$, the right hand side of (8.8) is a polynomial in α of degree twelve. Moreover, the term independent of α is $\mu^2(\mu^2 - 1) \neq 0$, and hence for each $\mu > \sqrt{2}$ there are at most twelve solutions. Thus the set of rationally dependent solutions is a countable union of finite sets, and is hence countable (and measure zero). \square

9. The induced map K and interval exchange maps: case B. From (7.18) the monodromy invariant of K in (7.17) is $\pi_M = \begin{pmatrix} 3 & 1 & 4 & 2 \end{pmatrix}$ which is again irreducible. Using (7.15) and (7.16) the length vector λ has

$$(9.1) \quad \lambda_1 = \tan \psi(1 - \tan \tilde{\theta}), \quad \lambda_2 = 1 - \tan \theta - \tan \psi, \quad \lambda_4 = 1 - \tan \psi,$$

and $\lambda_3 = 1 - \lambda_1 - \lambda_2 - \lambda_4$. Hence, with suitable rescaling, the length vector is rationally dependent if and only if there exist rationals d_1, d_2 and d_3 not all zero such that

$$d_1 \tan \psi + d_2 \tan \theta + d_3 \tan \psi \tan \tilde{\theta} = 1.$$

This is precisely the same three fundamental quantities, $\tan \psi, \tan \theta$ and $\tan \psi \tan \tilde{\theta}$, that appeared in the previous section (8.6) and hence by the same argument for

each $\mu > \sqrt{2}$ these are rationally independent except on a countable set of θ . This completes the minimality result:

LEMMA 9.1. *If $\mu > \sqrt{2}$ then for all but a countable set of $\theta \in (\hat{\theta}, \theta_{cc})$ all orbits of (7.7) are minimal.*

Putting Lemmas 8.2 and 9.1 together yields the result stated in the introduction: if $\mu > \sqrt{2}$ then except for a countable set of incidence angles all rays that are refracted at some stage have dense intersection with the boundary modulo the identification by rotations and integer translations through multiples of 2 and rotations by $\frac{\pi}{2}$. Of course, this implies that almost all rays are neither periodic nor periodic modulo integer translations and rotations by $\frac{\pi}{2}$, making it the analogue of the dichotomy of rigid rotations between irrational rotations (all orbits dense) and rational rotations (all orbits periodic).

10. The behaviour of orbits. In this section we will show how the induced map can be used to explain qualitative features of rays. With irrational rotations, nearby rationals give clues about the dynamics: for example if $x \rightarrow x + \alpha \pmod{1}$ with $\alpha = \frac{1}{2} + \epsilon$ is irrational and $|\epsilon| \ll 1$ then although the motion is dense on the circle it may be more helpful to know that the motion is close to periodic of period two, with a slow drift around the circle.

To illustrate the equivalent phenomenon for the light rays studied here we start by considering the special case

$$(10.1) \quad \tan \theta + \tan \psi = \frac{1}{4}, \quad \mu > \sqrt{2}$$

which is *not* covered by the results of previous sections as it does not satisfy the Keane condition. To see this note that since $\tan \theta + \tan \psi < 1$ and $\mu > \sqrt{2}$ (10.1) corresponds to case A of previous sections and the induced map is H defined in (7.7). Putting (10.1) and (7.7) together shows that $H(x) = x - \frac{1}{4}$ if $x \in [\frac{1}{4}, 1) = I_4$ and then (7.11), (7.12), (7.7) and (10.1) with $x_1 = (1 - \tan \tilde{\theta}) \tan \psi$ imply that $H(x_1) = \frac{3}{4} \in I_4$. Thus two iterations of $H(x) = x - \frac{1}{4}$ gives $H^3(x_1) = \frac{1}{4} = x_3$, which violates the Keane condition (8.5).

An elementary calculation shows that this occurs if $\tan \psi = \alpha$ is the root of the quartic equation

$$(10.2) \quad \alpha^4 - 2k\alpha^3 + (1 + k^2)\alpha^2 - \frac{2k\mu^2}{\mu^2 - 1}\alpha + k^2 \frac{\mu^2}{\mu^2 - 1} = 0$$

with $k = \frac{1}{4}$. Equations (7.7) and (10.1) imply that $H(x) = x + \frac{3}{4}$ if $x \in [\tan \psi, \frac{1}{4}) = I_3$. We have already established that $H(x) = x - \frac{1}{4}$ if $x \in I_4$. Hence all points in $[\tan \psi, \frac{1}{4})$ map to $[\frac{3}{4}, 1)$ and after three iterations in I_4 arrive back in $x < \frac{1}{4}$ at their starting point, i.e. they are periodic with period four. A similar calculation shows that H^4 maps $(0, \tan \psi)$ into itself as a rigid rotation with rotation number $\tan \tilde{\theta}$. Since $\tilde{\theta}$ varies continuously with μ this implies that for almost all μ the motion is dense restricted to appropriate intervals (irrational rotation number) and otherwise all orbits are periodic (rational rotation number).

Figure 6 shows three light rays computed numerically with (10.1) and $\mu = 1.5$. For this value numerical solution of (10.2) gives

$$(10.3) \quad \begin{aligned} \alpha = \tan \psi &\approx 0.1503743, & \psi &\approx 0.1492560, \\ \theta &\approx 0.0992981, & \tan \tilde{\theta} &\approx 0.8767620. \end{aligned}$$

In Figure 6a the motion looks periodic (modulo the symmetry of the square lattice) as we would expect from the initial value $x_0 \in I_3$. The dynamics of the induced map

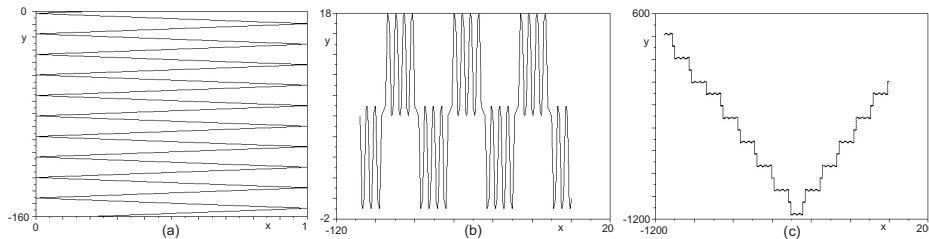


FIG. 6. Light rays with $\mu = 1.5$ and $y_0 = 0$, represented as in Figure 1. (a) $x_0 = 0.2$; $\theta_0 = 0.0992981$; 100 incidence events; (b) $x_0 = 0.1$; $\theta_0 = 0.0992981$; 300 incidence events; (c) $x_0 = 0.1$; $\theta_0 = 0.1$; 5000 incidence events.

is therefore periodic with one point in I_3 and one in I_4 . In I_4 , θ does not change and the motion corresponds to the behaviour labelled DD (Down-Down), i.e. the ray simply moves to the next incident plane in the direction it is moving (and this is repeated three times). In I_3 , $H(x) = D^2(x)$ with the first iteration DL (Down-Left, or Down-Right if θ is negative) indicating a change of direction and $|\theta| \rightarrow \frac{\pi}{2} - |\theta|$, and this is followed by a reflection which changes the direction again and restores the angle to $-\theta$. This explains the zig-zag motion observed, with net motion in one of the four directions perpendicular to the incidence planes, depending on the initial direction.

Figure 6b shows a ray with $x_0 = 0.1 \in (0, \tan \psi)$ – here every fourth iteration of H involves either LL or LD, and so the dynamics is more complicated, but easy enough to explain by interpreting the iteration (we do not do this for reasons of space). Finally Figure 6c shows a ray with $\theta = 0.1$ initially, showing long periods of motion close to the periodic orbit of Figure 6a, followed by circulation similar to Figure 6b.

11. Conclusion. Although Snell's Law (Descartes' Law) has been discussed in many contexts, we have been unable to find any description of the dynamics of rays in complicated media. This paper makes a first attempt to show the variety of behaviour that is possible in extended media. A chessboard configuration of materials was chosen for convenience, though other configurations would be interesting to pursue. The result is a description of the rays from a purely geometric point of view; a description that has much in common with the study of billiards.

For the chessboard we have established that the set of angles of incidence is finite, and in the simplest case we define an induced map that is an interval exchange map. This seems a particularly natural description given the reversible nature of the light rays and we conjecture that this is much more general than the particular context described here. This enables us to show that for all but a finite angles of incidence the ray is dense on the incident plane (modulo the symmetries of the chessboard).

For one of the special cases where the rays are not dense on the incident plane we describe some features of solutions; for example the existence of well-defined directions of motion. It would be interesting to determine whether this can be extended to other solutions and more general directions.

The laws of refraction and reflection can also be seen as determining paths that minimize the time of passage between two points. Thus the finite angles results show an interesting feature of solutions to these problems. This is another area that could be taken further.

This paper is a first step. We believe the refraction/reflection dynamics provides an interesting generalization of billiards, whilst the use of more complicated arrange-

ments of materials has potential application in the study of light in complex media – a topic of growing interest in the physics community.

REFERENCES

- [1] E.G. Altmann, J.C. Leitão and J. Viana Lopes (2012) Effect of noise in open chaotic billiards *Chaos* **22** 026114, doi:10.1063/1.3697408
- [2] A. Avila and G. Forni (2007) Weak mixing for interval exchange transformations and translation flows *Ann. Math.* **135** 637–664, doi: 10.4007/annals.2007.165.637 r.
- [3] P. Bachurin, K. Khanin, J. Marklof and A. Plakhov (2011) Perfect retroreflectors and billiard dynamics *J. Mod. Dyn.* **5** 33–48, doi:10.3934/jmd.2011.5.33
- [4] V.G. Baryakhtar, V.V. Yanovsky, S.V. Naydenov and A.V. Kurilo (2006) Chaos in composite billiards, *J. Exp. Theor. Phys.*, **103** 292–302, doi:10.1134/S1063776106080127
- [5] L.A. Bunimovich (1979) On the Ergodic Properties of Nowhere Dispersing Billiards *Comm. Math. Phys.* **65** 295–312.
- [6] V. Delecroix, P. Hubert and S. Lelièvre (2014) Diffusion for the periodic wind-tree model, *Annales scientifiques de l'ENS* **47** 1085–1110.
- [7] C.P. Dettmann and O. Georgiou (2012) Quantifying intermittency in the open drivebelt billiard *Chaos* **22** 02611, doi:10.1063/1.3685522
- [8] K. Frzcek and C. Ulcigrai (2014) Non-ergodic \mathbb{Z} -periodic billiards and infinite translation surfaces, *Invent. Math.* **197** 241–298.
- [9] P. Hubert, S. Lelièvre and S. Troubetzkoy (2011) The Ehrenfest wind-tree model: periodic directions, recurrence, diffusion *Journal für die reine und angewandte Mathematik* **2011** 223–244.
- [10] M. Keane (1975) Interval exchange transformations *Math. Z.* **141** 25–31.
- [11] A. Lopes and R. Markarian (1996) Open Billiards: Invariant and Conditionally Invariant Probabilities on Cantor Sets *SIAM J. Appl. Math.* **56** 651–680.
- [12] H. Mazur and S. Tabachnikov (2002) Rational Billiards and Flat Structures, in *Handbook of Dynamical Systems Vol. 1A* Eds. B. Hasselblatt and A. Katok, Elsevier.
- [13] M. Rees (1981) An alternative approach to the ergodic theory of measured foliations on surfaces *Ergod. Th. Dyn. Sys.* **1** 461–488.
- [14] J. Smillie (2000) The Dynamics of Billiard Flows in Rational Polygons of Dynamical Systems, in *Dynamical Systems, Ergodic Theory and Applications*, eds. L.A. Bunimovich & Y.G. Sinai, 360–382.
- [15] S. Tabachnikov (2005) *Geometry and Billiards*, Student Mathematical Library 30, AMS, Providence, RI.
- [16] W. A. Veech (1978) Interval exchange transformations, *J. Analyse Math.* **33** 222–272.
- [17] W. A. Veech (1986) The Teichmüller geodesic flow *Ann. Math.* **124** , 441–530.
- [18] M. Viana (2006) Ergodic Theory of Interval Exchange Maps *Rev. Mat. Complut.* **19** 7–100.
- [19] J.C. Yoccoz (2010) Interval exchange maps and translation surfaces, *Clay Math. Proc.* **10** 1–70.