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 sytems 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 rational billiards and can 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, amongst other things, 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. [11]. In two dimensions the study of billiards is often described by a pair (s, ψ) where s is some arclength parametrization of the boundary and ψ is the angle of incidence of the particle. Thus successive impacts of the particle on the boundary generate a sequence (s_n, ψ_n) where the n^{th} collision with the boundary is simply a function of the position and angle of the previous

*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.

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

collision. Viewed as a map this process has many nice properties depending on the nature of the boundary on which collisions take place, the most famous being the stadium, though boundaries with corners such as squares are also considered, though the behaviour on hitting a corner is not well-defined, so it is usual to ignore such collisions which are typically of zero measure anyway. 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 how many 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 [5, 7] (because many billiard systems are ergodic, almost all solutions do eventually leave), how does this change with the shape of the billiard table [3]?

It is equally possible to think of billiards as giving the intersections of the path of a ray of light with a mirrored surface [11]. 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 medium A to B say) the ray always passes through, whilst in the other direction there is a critical angle θ_c such that if $\theta > \theta_c$ then the ray is reflected. As the ratio of refractive indices tendss 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*. They describe the diffusion of light out of the central region and discuss 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 we replace the bounded property of their geometry with an infinite lattice by considering 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.

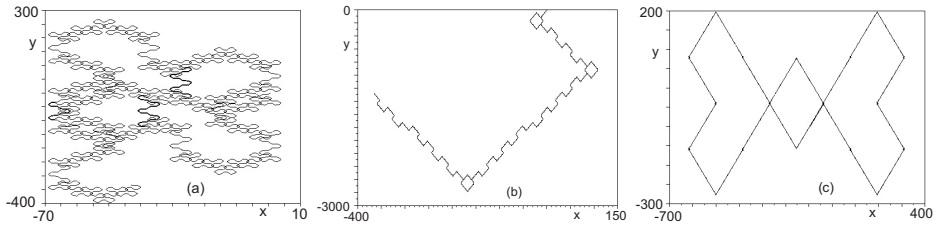

FIG. 1. *Rays showing 5000 refractions and reflections with μ defined above (2.1). (a) $x_0 = 0.6$; $\theta_0 = 0.1$; $\mu = 1.9$; (b) $x_0 = 0.6$; $\theta_0 = 0.01$; $\mu = 7$; and (c) $x_0 = 0.6$; $\theta_0 = 0.1$; $\mu = 7$.*

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. We also illustrate the action of this map. Some basic properties are derived 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 rays can be described by an induced map which is an *interval exchange map*.
- If $\mu > \sqrt{2}$ then the set of incidence angles that lead to rays which do not intersect the boundary (taken modulo rotations by $\frac{m\pi}{2}$ about the mid-point of a square and the invariant translations of the chessboard) at a dense set of points is countable.

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 $x \to x + \alpha \ (mod\ 1)$ that motion is dense on the circle except at a countable number of values of α (the rationals), where every point is periodic.

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 smaller than the index n_b in the material labelled B (think of W as air and B as water or glass). Let $\mu = n_b/n_w > 1$. 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

$$\mu \sin\theta = \sin\psi. \tag{2.1}$$

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 bouindary with angle ψ will be refracted using (2.1), so solutions are invertible. Solutions striking the boundary at corners are not well-defined and 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 at a given point x at an angle of incidence θ. 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, and the arclength of the boundary measured clockwise from $(0,0)$ will be $z\in[0,4)$. Let $x=z-\lfloor z\rfloor$ and take the angle of incidence of a ray about to pass out of the square to be in the interval $(-\frac{\pi}{2},\frac{\pi}{2})$ with positive angles in the direction of increasing z on each face.

Now a ray incident on the boundary with $z\geq 1$ and angle of incidence θ is equivalent by rotational symmetry through multiples of $\frac{\pi}{2}$ to a ray incident on the lower boundary at $z=x$. 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

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

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 square at x with angle of incidence $\theta>0$ will be reflected if $\theta>\theta_c$ where

$$\theta_c=\sin^{-1}(1/\mu) \tag{2.2}$$

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

2.1. Reflections. If $\theta>\theta_c$ and $\theta>\frac{\pi}{4}$ then by elementary geometry the reflected ray will hit the left hand face at a coordinate x' (in the sense of the previous paragraph), i.e. $z=3+x'$, where $\tan\theta=x/(1-x')$ and with angle of incidence $\frac{\pi}{2}-\theta$ and hence with coordinates (x',θ') where

$$x'=1-x/\tan\theta,\quad \theta'=\tfrac{\pi}{2}-\theta. \tag{2.3}$$

If $\theta_c<\frac{\pi}{4}$, i.e. if $\mu>\sqrt{2}$, then there is the additional possibility that angles of incidence lie in the interval $(\theta_c,\frac{\pi}{4})$. These will either hit the left face satisfying (2.3) or they are reflected back to the upper face and $z=2+x'$ where

$$x'=1-x-\tan\theta,\quad \theta'=\theta. \tag{2.4}$$

The boundary between these two cases occurs if the reflection is back to the top left hand corner, i.e. $x=\tan\theta$, in which case both (2.3) and (2.4) give the correct 'corner' $x'=0$ from (2.3) and $x'=1$ from (2.4). Thus to summarize, the reflection rules for a point with coordinates (x,θ) on the bottom face with $\theta>\theta_c$ are that if $0<x<\tan\theta$ then the next incidence with a white face is on the left face with position and angle given by (2.3), whilst if $\tan\theta<x<1$ then the next incidence is on the top face with coordinates given by (2.4).

2.2. Refractions: down-down. The refractions divide into four subcases as shown in Figure 2. In the first, 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'). After the first refraction, the angle of refraction (in the black square) is ψ where

$$\psi=\sin^{-1}(\mu\sin\theta), \tag{2.5}$$

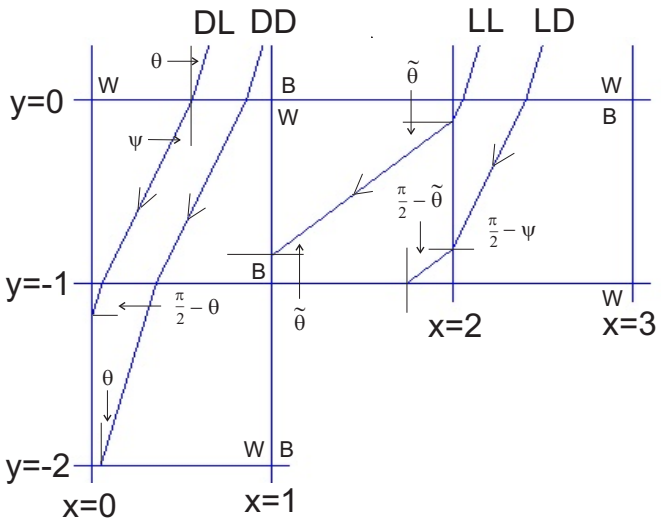

FIG. 2. *Different cases of refraction with $\mu = 1.512$ ($\mu > \sqrt{2}$) showing the geometry of different rays. All rays are incident on the white squares (W) with $y = 0$ at the same angle $\theta = 0.3$. In the notation of the text: DD – down-down; DL – down-left; LD – left-down; and LL – left-left. Also, the final incidence angles for DL and LL are treated as negative in the analysis as the ray is in the direction of decreasing arclength (see main text).*

and provided $\tan\psi < x$ then the intersection with the bottom of this back 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

$$x' = x - \tan\psi - \tan\theta, \quad \theta' = \theta. \tag{2.6}$$

2.3. Refractions: down-left. In Figure 2b 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 (z', θ') where $z' = 3 + x'$ with

$$x' = (x - \tan\psi)/\tan\theta, \quad \theta' = -(\tfrac{\pi}{2} - \theta). \tag{2.7}$$

2.4. Refractions: left-down. If $x < \tan\psi$ then the refrated 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 $\mu \sin\tilde{\theta} = \cos\psi$ or

$$\tilde{\theta} = \sin^{-1}(\tfrac{1}{\mu}\cos\psi). \tag{2.8}$$

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

$$x' = 1 - \left(1 - \frac{x}{\tan\psi}\right)/\tan\tilde{\theta}, \quad \theta' = \frac{\pi}{2} - \tilde{\theta}. \tag{2.9}$$

2.5. 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 2d with coordinate $z' = 3 + x'$ where

$$x' = (x/\tan\psi) + \tan\tilde\theta, \quad \theta' = -\tilde\theta. \tag{2.10}$$

3. A factorized return map. In each of the cases described in the previous subsections the light ray was incident on the bottom of a white square with a positive angle of incidence. By a sequence of reflections and rotations this information makes it possible to deduce the next internal incidence in a white square from any initial incident ray in a white square. Conversely we may identify incidence at $(x, -\theta)$ with incidence at $(1 - x, \theta)$ and the different faces by rotation. In this way the rules of the four refractions become, using the abbreviation DD for 'down-down', DL for 'down-left' etc:

$$\begin{array}{ll} DD & x' = x - \tan\psi - \tan\theta \qquad\quad \theta' = \theta \\ DL & x' = 1 - (x - \tan\psi)/\tan\theta \qquad \theta' = \frac{\pi}{2} - \theta \\ LD & x' = 1 - (1 - (x/\tan\psi))/\tan\tilde\theta \quad \theta' = \frac{\pi}{2} - \tilde\theta \\ LL & x' = 1 - \tan\tilde\theta - (x/\tan\psi) \qquad \theta' = \tilde\theta \end{array} \tag{3.1}$$

with ψ and $\tilde\theta$ defined in (2.5) and (2.8). These four maps have their domains in different part of the phase space. For DD the map is defined on

$$DD = \{(x,\theta) \mid \tan\psi + \tan\theta < x < 1,\ 0 < \theta < \hat\theta\} \tag{3.2}$$

(with the obvious abuse of notation) where if $\hat\psi$ is defined in terms of $\hat\theta$ via (2.5), i.e. $\hat\psi$ is the refracted ray in B if the angle of incidence from W is $\hat\theta$, then

$$\tan\hat\psi + \tan\hat\theta = 1. \tag{3.3}$$

For DL

$$DL = \{(x,\theta) \mid \tan\psi < x < \min(1, \tan\psi + \tan\theta),\ 0 < \theta < \theta_{cc}\}, \tag{3.4}$$

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

$$\theta_{cc} = \sin^{-1}(1/(\mu\sqrt{2})), \quad \psi_{cc} = \frac{\pi}{4}, \quad \tilde\theta_{cc} = \theta_{cc}. \tag{3.5}$$

The latter follows because if the refracted angle is $\frac{\pi}{4}$ an 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

$$LD = \{(x,\theta) \mid (1 - \tan\tilde\theta)\tan\psi < x < \min(1, \tan\psi),\ \theta'_\circ < \theta < \theta_\circ\} \tag{3.6}$$

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

$$(1 - \tan\tilde\theta_\circ)\tan\psi_\circ = 1; \tag{3.7}$$

and $\psi_\circ$ and $\tilde{\theta}_\circ$ are defined from $\theta_\circ$ by (2.5) and (2.8). The minimum in (3.6) equals 1 in $(\theta_{cc}, \theta_\circ)$ and $\tan\psi$ below this interval.

Finally, LL is defined by

$$LL = \{(x,\theta) \mid 0 < x < \min(1, (1-\tan\tilde{\theta})\tan\psi),\ 0 < \theta < \theta_c\} \tag{3.8}$$

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

Now consider the reflections, which occur if $\theta > \theta_c = \sin^{-1}(1/\mu)$. Following a reflection either θ is unchanged or it is $\frac{\pi}{2} - \theta$. So if both θ and $\frac{\pi}{2} - \theta$ are greater than θ_c the rays are always reflected and never leave the initial box. Thus if $\mu > \sqrt{2}$ and $\theta_c < \theta < \frac{\pi}{2} - \theta_c$, an inequality which is only possible if $\mu > \sqrt{2}$, any ray will be perpetually reflected, whilst if $\mu > \sqrt{2}$ and $\theta > \frac{\pi}{2} - \theta_c = \cos^{-1}(1/\mu)$ then after one reflection the new angle is $\frac{\pi}{2} - \theta < \theta_c$ and so every reflection is followed by a refraction. Since the study of reflections (classical billiards) is well-understood the case of perpetual reflections will not be analyzed further here and so only equation (2.3) will be used.

If $\mu < \sqrt{2}$ then the purely reflected states no longer exist and every reflection is followed by a reflection anyway, so once again there is no need to consider reflections followed by reflections as in (2.4). Thus if $\theta > \theta_c$ and, if $\mu > \sqrt{2}$, $\theta > \frac{\pi}{2} - \theta_c$, then the reflective map from (2.3)

$$RR \quad x' = 1 - x/\tan\theta, \quad \theta' = \tfrac{\pi}{2} - \theta \tag{3.9}$$

needs to be added to the map (3.1), together with the reflected-refracted domain RR defined by

$$RR = \{(x,\theta) \mid 0 < x < 1,\ \max(\theta_c, \frac{\pi}{2} - \theta_c) < \theta < \frac{\pi}{2}\} \tag{3.10}$$

Note that the return maps (3.1) and (3.9) have been obtained by factorizing out the symmetries of the chessboard (generated by rotations through $\frac{\pi}{2}$ and reflections in half interger horizontal and vertical lines and so the interpretation of solutions in terms of directions travelled and their 'true' positions are not transparently obvious from solutions of the factorized map, on the other hand the map is considerably easier to analyze than the full map and we will develop ways to extract information in section 8. For the moment, the geometry of images of the different regions on which the factorized map is defined will be the focus of the remainder of this section. It turns out that the maps are (obviously) piecewise affine, and that the images are disjoint (reflecting the reversibility of light rays) and the closure of the images is the whole space,

$$X = (0,1) \times (0, \frac{\pi}{2}). \tag{3.11}$$

Let F denote the factorized return on X defined by the refraction maps (3.1), and the two reflection maps (3.9) and (2.4). Note that in all cases where F is defined, horizontal lines of constant θ are mapped to horizontal lines. This remark will be crucial in sectionsect:lines.

One further result will be useful.

LEMMA 3.1. *With the notation above, i.e. (3.3), (3.5) and (3.7) with the tilde operation defined by (2.8),*

$$\tilde{\theta}_{cc} = \theta_{cc}, \quad \tilde{\theta}_\circ = \hat{\theta}, \quad \tilde{\hat{\theta}} = \theta_\circ, \tag{3.12}$$

and $\tilde{0} = \theta_c$, $\tilde{\theta}_c = 0$.

Proof: Since $\psi_c = \frac{\pi}{4}$, $\tilde{\psi}_c = \frac{\pi}{2} - \psi_c = \frac{\pi}{4}$ and hence both θ_c and $\tilde{\theta}_c$ have the same angle of refraction and are hence equal.

$\hat{\theta}$ is defined by (3.3) and $\mu \sin\hat{\theta} = \sin\hat{\psi}$. By definition, $\mu \sin\tilde{\hat{\theta}} = \cos\hat{\theta}$. Thus (3.3) implies that $1 = \tan\hat{\theta} + \tan\hat{\theta}$ and recall that the definition of $\theta_\circ$ is that $(1 - \tan\tilde{\theta}_\circ)\tan\psi_\circ = 1$ from (3.7).

Since

$$\tilde{\hat{\psi}} = \frac{\pi}{2} - \hat{\psi}$$

by definition, the definition of $\hat{\theta}$ can be written as $1 - \tan\hat{\theta} = \tan\hat{\psi} = 1/\tan\tilde{\hat{\psi}}$. Thus $\hat{\theta}$ satisfies the criterion for $\tilde{\theta}_\circ$. Since $\tilde{\tilde{\theta}} = \theta$ the third equality holds. The final part is straightforward from the definitions. □

3.1. Geometry of images, $\mu > \sqrt{2}$. If $\mu > \sqrt{2}$ then

$$PR = \{(x,\theta) \mid 0 < x < 1,\ \theta_c < \theta < \frac{\pi}{2} - \theta_c\} \tag{3.13}$$

is non-empty and $F(PR) = PR$, so any inital condition in PR is reflected for all time; PR stands for 'pure reflections'. On the other hand, an initial point in RR defined in (3.10) is reflected and its image is refracted. Note that $\mu > \sqrt{2}$ implies $\max(\theta_c, \frac{\pi}{2} - \theta_c) = \frac{\pi}{2} - \theta_c$.

As $\theta \to \frac{\pi}{2}$ from below then by (3.9), $x' \to 1$ ($x \neq 0$) and $\theta' \to 0$, so the limiting image of the line $\theta = \frac{\pi}{2}$ is singular; it is the point $(1,0)$. Similarly, as $\theta \to \frac{\pi}{2} - \theta_c$ from above then $\theta' \to \theta_c$ from below, and as $x \to 0$ the image $x' \to 1$, so the image of the line $0 < x < 1$, $\theta = \frac{\pi}{2} - \theta_c$ is the line $\tan\theta_c < x' < 1$, $\theta' = \theta_c$. Note that from the definition (2.2), $\tan\theta_c = 1/\sqrt{\mu^2 - 1}$.

Thus by continuity of F in RR, $F(RR)$ is an open region whose boundary consists of three curves with the 'topmost' on the line $\theta = \theta_c$ and the other two being curves connecting $(\tan\theta_c, \theta_c)$ to $(1,0)$ and $(1,\theta_c)$ to $(1,0)$ respectively. The image of $x = 0$ in RR is clearly $x' = 1$, so the latter curve is simply the straight line $x = 1$. From (3.9) the former is

$$\{(x',\theta') \mid x' = 1 - (1/\tan\theta),\ \theta' = \frac{\pi}{2} - \theta,\ \frac{\pi}{2} - \theta_c < \theta < \frac{\pi}{2}\}$$

which is more conveniently written as

$$\{(x,\theta) \mid x = 1 - \tan\theta,\ 0 < \theta < \theta_c\} \tag{3.14}$$

by reparametrising using θ' with the identity $\tan\theta\tan(\frac{\pi}{2} - \theta) = 1$. Numerical evaluation of this region is shown in Figure 3a, which also shows the images of the four other regions whose images need to be established. Thus $F(RR)$ is the generalized triangle with sides $x = 1$, part of $\theta = \theta_c$, and (3.14).

We now consider the four remaining regions. DD, defined by (3.2) is a three sided region bounded by $x = 0$, $x = \tan\psi + \tan\theta$ and the line $x = 1$ with $0 < \theta < \hat{\theta}$, where $1 = \tan\hat{\psi} + \tan\hat{\theta}$, with $\hat{\psi}$ defined from $\hat{\theta}$ by (2.5). In the limit $\theta \to 0$, (2.6) is the identity, so the lower boundary is unchanged. By definition the boundary curve with $x = \tan\psi + \tan\theta$ is mapped to $x' = 0$ with θ unchanged whilst the line segment $x = 1$, $0 < \theta < \hat{\theta}$, is mapped to the curve

$$x' = 1 - \tan\theta - \tan\psi, \quad 0 < \theta < \hat{\theta} \tag{3.15}$$

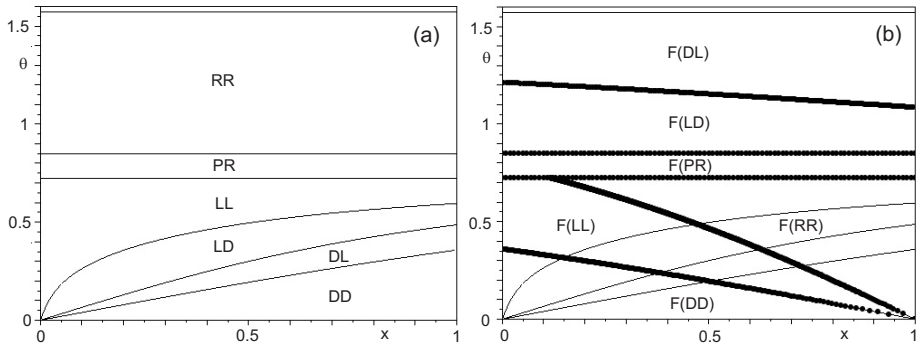

FIG. 3. *Regions and their images in the (x,θ)-plane if $\mu = 1.512$, i.e. $\mu > \sqrt{2}$. (a) Basic regions; (b) images of the basic regions.*

which stretches from $(0,\hat{\theta})$ to $(1,0)$ making $F(DD)$ a triangular region as shown.

The region DL is defined in (3.4), so it is a triangular region, with the factorized return map defined by (3.1). It has boundary

$$\begin{array}{ll} x = \tan\psi, & 0 < \theta < \theta_{cc} \\ x = \tan\psi + \tan\theta, & 0 < \theta < \hat{\theta} \\ x = 1, & \hat{\theta} < \theta < \theta_{cc}. \end{array} \tag{3.16}$$

By direct calculation the image of the first of these has $x' = 1$ and $\theta' = \frac{\pi}{2} - \theta$ and hence it is the set

$$\{(x,\theta) \mid x = 1,\ \frac{\pi}{2} - \theta_{cc} < \theta < \frac{\pi}{2}\}. \tag{3.17}$$

The second has $x' = 0$ and so its image is the set

$$\{(x,\theta) \mid x = 0,\ \frac{\pi}{2} - \hat{\theta} < \theta < \frac{\pi}{2}\}. \tag{3.18}$$

The image of $x = 1$ has $x' = 1 - (1 - \tan\psi)/\tan\theta$ with $\theta' = \frac{\pi}{2} - \theta$ and so the image of this line is

$$\{(x',\theta') \mid x' = 1 - (1 - \tan\psi)/\tan\theta,\ \theta' = \frac{\pi}{2} - \theta,\ \hat{\theta} < \theta < \theta_{cc}\}. \tag{3.19}$$

and note that this latter curve lies between $\frac{\pi}{2} - \theta_{cc}$ and $\frac{\pi}{2} - \hat{\theta}$. Moreover, the image of smaller and smaller horizontal lines as $\theta \to 0$ in DL map to horizontal lines that stretch from 0 to 1 near $\frac{\pi}{2}$, so this is a singular limit.

The region LD is bounded by three curves:

$$\begin{array}{ll} x = \tan\psi, & 0 < \theta < \theta_{cc} \\ x = (1 - \tan\tilde{\theta})\tan\psi, & 0 < \theta < \theta_\circ \\ x = 1, & \theta_{cc} < \theta < \theta_\circ. \end{array} \tag{3.20}$$

Using equation LD of (3.1) the image of the first has $x' = 1$ and $\theta' = \frac{\pi}{2} - \tilde{\theta}$. Since $\tilde{\theta}_{cc} = \theta_{cc}$ and $\tilde{0} = \theta_c$ this gives the set

$$\{(x,\theta) \mid x = 1,\ \frac{\pi}{2} - \theta_c < \theta < \frac{\pi}{2} - \theta_{cc}\}. \tag{3.21}$$

Similarly a direct calulation and the fact that $\tilde{0} = \theta_c$ and $\tilde{\theta}_\circ = \hat{\theta}$ gives that the image of the second is

$$\{(x,\theta) \mid x = 0,\ \frac{\pi}{2} - \theta_c < \theta < \frac{\pi}{2} - \hat{\theta}\}. \tag{3.22}$$

Finally, if $x = 1$ then $x' = 1 - (1 - (1/\tan\psi))/\tan\tilde{\theta})$ with $\theta' = \frac{\pi}{2} - \tilde{\theta}$. We will write this in terms of $\tilde{\theta}$, with boundaries $\tilde{\theta}_\circ = \hat{\theta}$ and $\tilde{\theta}_{cc} = \theta_{cc}$. Now, if angle of incidence θ produces and angle of refraction ψ then $\tilde{\theta}$ produces angle of refraction $\tilde{\psi} = \frac{\pi}{2} - \psi$ and so since $\tan\psi\tan(\frac{\pi}{2} - \psi) = 1$ the set can be written as

$$\{(x',\theta') \mid x' = 1 - (1 - \tan\tilde{\psi})/\tan\tilde{\theta},\ \theta' = \frac{\pi}{2} - \tilde{\theta},\ \hat{\theta} < \tilde{\theta} < \theta_{cc}\}. \tag{3.23}$$

This is of course precisely the same set as (3.19).

The boundary of LL has four parts:

$$\begin{array}{ll} x = 0, & 0 < \theta < \theta_c \\ x = (1 - \tan\tilde{\theta})\tan\psi, & 0 < \theta < \theta_\circ \\ x = 1, & \theta_\circ < \theta < \theta_c \\ \theta = \theta_c, & 0 < x < 1. \end{array} \tag{3.24}$$

The image of the last of these has $x' = 1$ and $\theta' = \tilde{\theta}_c = 0$, so the entire line maps to $(1, 0)$.

If $x = 0$, $0 < \theta < \theta_c$ then $x' = 1 - \tan\tilde{\theta}$ for $\tilde{\theta}$ between $\tilde{0} = \theta_c$ and $\tilde{\theta}_c = 0$, so it is the set

$$\{(x,\theta) \mid x = 1 - \tan\theta,\ 0 < \theta < \theta_c\} \tag{3.25}$$

which is precisely the same as the image of the boundary of $F(RR)$ in (3.14).

If $x = 1$ with $\theta_\circ < \theta < \theta_c$ then the image is $x' = 1 - \tan\tilde{\theta} - 1/(\tan\psi)$ with $\theta' = \frac{\pi}{2} - \tilde{\theta}$. Thus parametrizing solutions via the angle $\tilde{\theta}$ and using the relation $\tan\psi\tan\tilde{\psi} = 1$ as before, with $\tilde{\theta}_c = 0$ and $\tilde{\theta}_\circ = \hat{\theta}$ we find that the solutions lie on the set

$$\{(x,\tilde{\theta}) \mid x = 1 - \tan\tilde{\theta} - \tan\tilde{\psi},\ 0 < \tilde{\theta} < \hat{\theta}\} \tag{3.26}$$

which is precisely the diagonal boundary of $F(DD)$ derived earlier.

The final boundary is $x = \tan\tilde{\theta})\tan\psi$ with $0 < \theta < \theta_\circ$. Substituting into equation LL of (3.1) gives $x' = 0$ and $\theta' = \tilde{\theta}$, and since $\hat{\theta} < \tilde{\theta} < \theta_c$ this is the last part of the triangular region marked $F(LL)$ in Figure fig:regionsa.

Thus the images of the regions are disjoint and the union of they form a partition of the phase space.

3.2. Geometry of images, $1 < \mu < \sqrt{2}$. In this case the region PR does not exist as $\theta_c > \frac{\pi}{2} - \theta_c$, and the only other difference is that the curve separating LD and LL does not pass though the origin, but strikes the line $x = 0$ at a positive value of θ, see the remarks below (3.6). Apart from this the basic geometry and arrangement of the images is unchanged as illustrated in Figure 4.

4. 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

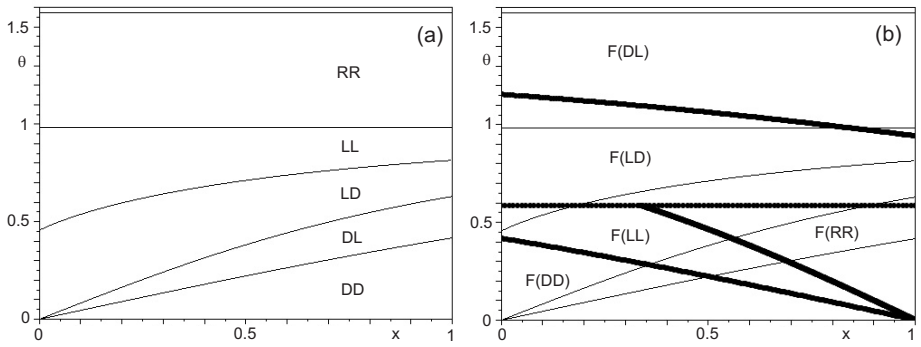

FIG. 4. *Regions and their images in the (x,θ)-plane for $\mu = 1.2$, i.e. $\mu < \sqrt{2}$. (a) Basic regions; (b) images of the basic regions.*

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 th 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 $\mu > 1$ solutions have a finite number of possible angles that can be generated in this way, and hence that the θ behaviour of the factorized map always involves just a 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.

LEMMA 4.1. *If $\mu > \mu_n$,*

$$\mu_n = \sqrt{\frac{n+1}{n}} \tag{4.1}$$

$n = 1, 2, \ldots$, then given any initial value of θ the orbit of (x,θ) takes at most $2n+2$ different values of the angle variable under iteration.

Proof: Given an intial $\theta_0 < \theta_c$ then by (3.1) 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

$$\begin{aligned} \mu \sin\theta_0 &= \sin\psi_0 \\ \mu \sin\tilde{\theta}_0 &= \cos\psi_0 = \sin(\tfrac{\pi}{2} - \psi_0) \end{aligned} \tag{4.2}$$

and $\theta_0 < \theta_c$ implies that $\tilde{\theta}_0 > \frac{\pi}{2} - \theta_c$, $\frac{\pi}{2} - \theta_0 > \frac{\pi}{2} - \theta_c$, and $\frac{\pi}{2} - \tilde{\theta}_0 < \theta_c$.

Thus θ_0 and $\tilde{\theta}_0$ are related through a refracted ray which is incident within a black square on perpendicular sides. The only way that new incidence angles can be created is via a similar process with the other two angles.

Start with $\theta_0' = \frac{\pi}{2} - \tilde{\theta}_0$ which, like θ_0 is less than θ_c. This can create a new angle of incidence if there exists ψ_1 and $\tilde{\theta}_1$ such that

$$\begin{aligned} \mu \sin\theta_1 &= \sin\psi_1 \\ \mu \sin\tilde{\theta}_1 &= \cos\psi_1. \end{aligned} \tag{4.3}$$

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 (4.3) implies

$$\begin{aligned}\sin\psi_1 &= \mu\sin\theta_1 = \mu\sin(\tfrac{\pi}{2}-\tilde\theta_0) = \mu\cos\tilde\theta_0\\ &= \sqrt{\mu^2-\mu^2\sin^2\tilde\theta_0} = \sqrt{\mu^2-\cos^2\psi_0}\\ &= \sqrt{\mu^2-1+\sin^2\psi_0}\end{aligned} \tag{4.4}$$

and so defining

$$u_0 = \sin\psi_0, \quad u_1 = \sin\psi_1 \tag{4.5}$$

this implies that

$$u_1 = \sqrt{\mu^2-1+u_0^2} \tag{4.6}$$

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, like $\tilde\theta_0$, is greater than $\frac{\pi}{2} - \theta_c$. The equivalent of (4.2) for the new angle θ_{-1} uses the angle ψ_{-1} defined by

$$\begin{aligned}\mu\sin\theta_{-1} &= \sin\psi_{-1}\\ \mu\sin\tilde\theta_{-1} &= \cos\psi_{-1}.\end{aligned} \tag{4.7}$$

Thus

$$\begin{aligned}\sin\tilde\psi_{-1} &= \sqrt{1-\cos^2\psi_{-1}} = \sqrt{1-\mu^2\sin^2\tilde\theta_{-1}} - \sqrt{1-\mu^2\sin^2(\tfrac{pi}{2}-\theta_0)}\\ &= \sqrt{1-\mu^2\cos^2\theta_0} = \sqrt{1-\mu^2+\mu^2\sin^2\theta_0}\\ &= \sqrt{1-\mu^2+\sin^2\psi_0}.\end{aligned} \tag{4.8}$$

Thus setting

$$u_{-1} = \sin\psi_{-1} \tag{4.9}$$

with u_0 as in (4.5),

$$u_{-1} = \sqrt{1-\mu^2+u_0^2} \tag{4.10}$$

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 (4.6) and (4.10). However, G is defined by

$$G(u) = \sqrt{\mu^2-1+u^2} \tag{4.11}$$

so that (4.6) is $u_1 = G(u_0)$, then

$$G^{-1}(u) = \sqrt{1-\mu^2+u^2} \tag{4.12}$$

and so (4.10) 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.

$$\mathcal{G} = \{u_k \in [0,1] \mid u_k = G^{k-1}(u_0),\ k \in \mathbb{Z}\}. \tag{4.13}$$

Each element of $\mathcal{G}$ is 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

$$2 + 2g \tag{4.14}$$

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 $\frac{\pi}{2} - \theta_{-m}$ 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

$$G^n(0) = \sqrt{n(\mu^2 - 1)} \tag{4.15}$$

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 $\mu \to 0$ 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). □

5. Induced maps for $\mu > \sqrt{2}$. If $\mu > \sqrt{2}$ then each orbit generates at most four incidence angles, and so the induced map (3.1) for an initial condition $(., \theta)$ 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 (3.1) if we ignore those angles that give the classic billiards of pure reflections. Moreover, without loss of generality the initial θ may be taken to be less than θ_c (as if it is greater then $\frac{\pi}{2} - \theta_c$ it is reflected and then refracted with incident angle θ) and since $\mu > \sqrt{2}$ (4.2) implies that both θ and $\tilde{\theta}$ are less than $\frac{\pi}{4}$ and hence less than θ_c, i.e. both $\frac{\pi}{2} - \theta$ and $\frac{\pi}{2} - \tilde{\theta}$ are greater than $\frac{\pi}{2} - \theta_c$, and on reflection lead to θ or $\tilde{\theta}$ respectively. Choose the labelling so that $\theta < \tilde{\theta}$, in which case it is quite easy to show that $\theta < \theta_{cc}$ (if $\theta = \theta_{cc}$ then $\theta = \tilde{\theta}$, and 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 (2.4) and (3.1) induce a one dimensional map D with discontinuities on the interval $[0, 4]$ where (choosing the easier induced maps first:

$$\begin{aligned} D(x) &= 1 - (x-3)/\tan(\tfrac{\pi}{2} - \theta) \\ &= 1 - (x-3)\tan\theta \quad \text{if } x \in [3, 4] \end{aligned} \tag{5.1}$$

and similarly

$$D(x) = 2 - (x-2)\tan\tilde{\theta} \quad \text{if } x \in [2, 3] \tag{5.2}$$

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,

$$D(x) = \begin{cases} x - \tan\psi - \tan\theta & \text{if } x \in (\tan\psi + \tan\theta, 1) \quad (DD) \\ 4 - (x - \tan\psi)\cot\theta & \text{if } x \in (\tan\psi, \tan\psi + \tan\theta) \quad (DL) \\ 3 - (1 - x\cot\psi)\cot\tilde{\theta} & \text{if } x \in ((1 - \tan\tilde{\theta})\tan\psi, \tan\psi) \quad (LD) \\ 2 - \tan\tilde{\theta} - x\cot\psi & \text{if } x \in (0, (1 - \tan\tilde{\theta})\tan\psi) \quad (LL) \end{cases} \tag{5.3}$$

where the shifts have been chosen to ensure that the images lie in the intervals with the appropriate angle variable consistent with (3.1). 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).

Similarly 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 and 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. $\theta < \hat{\theta}$ there is no zero in $(1, 2)$ and the only branch of the map that can be applied is LL, so

$$D(x) = 1 - \tan\theta - (x-1)\tan\psi \quad \text{if } x \in (1,2) \quad (LL) \tag{5.4}$$

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

$$D(x) = \begin{cases} 1 - \tan\theta - (x-1)\tan\psi & \text{if } x \in (1, x') \quad (LL) \\ 4 - (1 - (x-1)\tan\psi)\cot\theta & \text{if } x \in (x', 2) \quad (LD) \end{cases} \tag{5.5}$$

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

5.1. Case A. If $0 < \theta < \hat{\theta}$ where $\tan\hat{\theta} + \tan\hat{\psi} = 1$, the return map has seven branches defined by (5.1), (5.2), all four branches of (5.3) and (5.4). 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

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

Let us establish a little more of the geometry of the map on each of these intervals as shown in Fig 5. On I_4, the first equation of (5.3) holds and

$$D(I_4) = (0, 1 - \tan\psi - \tan\theta). \tag{5.6}$$

Now note that using (5.1), $D(3, 4) = (1 - \tan\theta, 1)$, so if $x \in I_3$, i.e. $\tan\psi < x < \tan\psi + \tan\theta$ then the map labelled DL in (5.3) 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

$$D^2(x) = x + 1 - \tan\psi - \tan\theta, \quad x \in (\tan\psi, \tan\psi + \tan\theta). \tag{5.7}$$

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

$$D^2(I_3) = (1 - \tan\theta, 1). \tag{5.8}$$

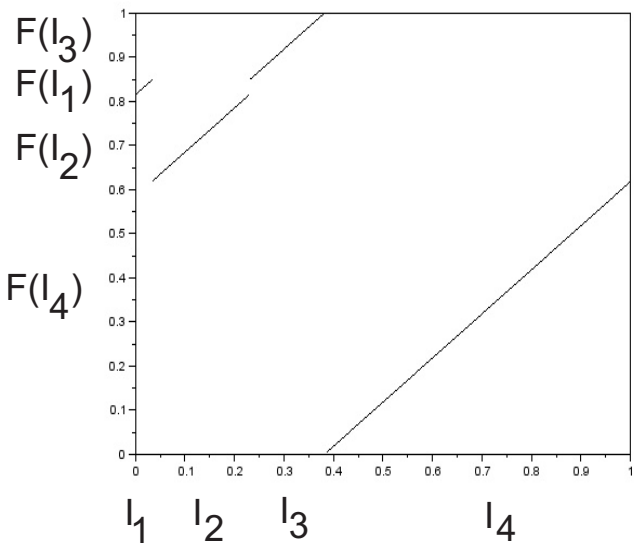

FIG. 5. *Induced map H with $\mu = 1.5$ and $\theta = 0.15$.*

Now consider $x \in I_2$, with D defined by the third equation of (5.3) 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 (5.2) 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 (5.4) (with slope $\tan\psi$) and so the image lies in $(0, 1)$ and D^3 has slope 1. Moreover,

$$D^3(I_2) = (1 - \tan\theta - \tan\psi, 1 - \tan\theta - \tan\psi + \tan\tilde{\theta}\tan\psi) \tag{5.9}$$

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

$$D^3(x) = x + 1 - 2\tan\psi - \tan\theta + \tan\tilde{\theta}\tan\psi. \tag{5.10}$$

Finally consider $x \in I_1$ which is the region labelled LL in (5.3), 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 (5.4) applies and so the second iterate of the right hand end point of LL in$(0, 1)$ is the image of $x = 1$ under (5.4) 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 (5.3)) and $\tan\psi$ from (5.4), i.e. it is 1 again and

$$D^2(I_1) = (1 - \tan\theta - \tan\psi + \tan\tilde{\theta}\tan\psi, 1 - \tan\theta) \tag{5.11}$$

and if $x \in I_1$ then

$$D^2(x) = x + 1 - \tan\theta - (1 - \tan\tilde{\theta})\tan\psi. \tag{5.12}$$

The last four paragraphs show that the induced map

$$H(x) = \begin{cases} D^2(x) & \text{if } x \in I_1 \\ D^3(x) & \text{if } x \in I_2 \\ D^2(x) & \text{if } x \in I_3 \\ D(x) & \text{if } x \in I_4 \end{cases} \tag{5.13}$$

is an *interval exchange map* [14] provided we extend it to the whole of the interval $[0,1)$ by defining the values at x_n, $n = 0,1,2,3$ using continuity from the right, which is equivalent to choosing the value at the discontinuities, a set of measure zero, conveniently. Interval exchange maps have been studied in many contexts since the mid-1970s [2, 6, 12] 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 [9, 13] that we return to in the conclusion.

5.2. Case B. If $\hat{\theta} < \theta < \theta_{cc}$ where $\tan\hat{\theta} + \tan\hat{\psi} = 1$, the return map has seven brances defined by (5.1), (5.2), all four branches of (5.3) and (5.4). We will refer to this as case B. In this case the return map is defined by (5.1), (5.2), the latter three branches of (5.3) and the two branches of (5.5), separated by x' where $(x'-1)\tan\psi = 1 - \tan\theta$. Define

$$y_0 = 0, \quad y_1 = (1-\tan\tilde{\theta})\tan\psi, \quad y_2 = Y, \quad y_3 = \tan\psi, \quad y_4 = 1. \tag{5.14}$$

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

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

$$(1-\tan\tilde{\theta})\tan\psi < 1 - \tan\theta$$

or

$$\tan\psi < \frac{1-\tan\theta}{1-\tan\tilde{\theta}}$$

(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 (5.2) 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 (5.14) 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

$$Y = \tan\psi(1 - \tan\tilde{\theta}) + \tan\theta + \tan\psi - 1. \tag{5.15}$$

With these preliminaries the calculations are almost identical to those in Case A, except the induced map is

$$K(x) = \begin{cases} D^2(x) & \text{if } x \in J_1 \\ D^4(x) & \text{if } x \in J_2 \\ D^3(x) & \text{if } x \in J_3 \\ D^2(x) & \text{if } x \in J_4 \end{cases} \tag{5.16}$$

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

$$K(J_3) \quad K(J_1) \quad K(J_4) \quad K(J_2) \tag{5.17}$$

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

6. The induced map H and interval exchange maps: case A. An interval exchange map does exactly what its name suggests. The interval $[0, 1)$ is partitioned into $N > 1$ subintervals $I_n = [x_{n-1}, x_n)$, $n = 1, \ldots, N$, 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 (5.13), 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

$$\pi_M = (\ 4\ \ 2\ \ 1\ \ 3\ ) \tag{6.1}$$

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 (6.1) is irreducible. Rather than use the translation vector, it is more usual to use the length vector $\lambda \in \mathbb{R}^N$, with $\lambda_n = x_n - x_{n-1}$, to define the interval exchange map. In the case of H, where $N = 4$,

$$\lambda = ((1 - \tan\tilde{\theta})\tan\psi, \tan\tilde{\theta}\tan\psi, \tan\theta, 1 - \tan\theta - \tan\psi). \tag{6.2}$$

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

$$\sum_1^N b_k\lambda_k = 0. \tag{6.3}$$

Note that since the sum of the λ_k equals to one, one of the components, λ_r say, can be eliminated and so (6.3) is equivalent to the existence of $r \in \{1, \ldots, N\}$ and $c \in \mathbb{Q}^{N-1}$, $c \neq 0$, such that

$$\sum_{k \neq r} c_k\lambda_k = 1. \tag{6.4}$$

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 wil be rationally independent. Finally we say that an interval exchange map f satisfies the *Keane condition* if for all $n \in \{0, \ldots, N-1\}$ and $r \geq 1$

$$f^r(x_i) \notin \{x_1, \ldots x_{N-1}\}. \tag{6.5}$$

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 (6.5).

A map $f : I \to I$ 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 [14] for example.

THEOREM 6.1. *Suppose $f : [0,1) \to [0,1)$ 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 (6.5).
(b) f is minimal if and only if f satisfies the Keane condition.

This also implies that almost all interval exchange maps are minimal [14], although of course our maps H are not generic so this is harder to apply directly. Indeed,from (6.2) and (6.4) 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

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

or equivalently rationals c_1, c_2 and c_3 at least one of which is non-zero such that

$$c_1\tan\psi + c_2\tan\tilde{\theta}\tan\psi + c_3\tan\theta = 1. \tag{6.6}$$

Using the parametrization $\alpha = \tan\psi$ this becomes

$$c_1\alpha + \frac{c_2\alpha}{\sqrt{\mu^2 - 1 + \mu^2\alpha^2}} + \frac{c_3\alpha}{\sqrt{(\mu^2-1)\alpha^2 + \mu^2}} = 1. \tag{6.7}$$

This will enable us to prove the following lemma.

LEMMA 6.2. *If $\mu > \sqrt{2}$ then for all but a countable set of $\theta \in (0, \hat{\theta})$ all orbits of (5.13) are minimal.*

Proof: By Theorem 6.1 we need only show that the values of α which satisfy (6.7) for rational c_k not all identically zero has measure zero.

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 (6.7) then it also satisfies

$$\begin{aligned} 0 = \quad & -4c_2^2c_3^2\alpha^4 p_1(\alpha)p_2(\alpha) \\ & +\big[(1-c_1\alpha)^2 p_1(\alpha)p_2(\alpha) - c_2\alpha^2 p_2(\alpha) - c_3\alpha^2 p_1(\alpha)\big]^2 = 0. \end{aligned} \tag{6.8}$$

For each of the countable set of rationals $(c_1, c_2, c_3) \neq (0,0,0)$, the right hand side of (6.8) is a polynomial in α of degree twelve. Morover, 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). □

7. The induced map K and interval exchange maps: case B. From (5.17) the monodromy invariant of K in (5.16) is

$$\pi_M = (\ 3\ \ 1\ \ 4\ \ 2\ ) \tag{7.1}$$

which is again irreducible. Using (5.14) and (5.15) the length vector λ has

$$\lambda_1 = \tan\psi(1-\tan\tilde{\theta}), \quad \lambda_2 = 1-\tan\theta-\tan\psi, \quad \lambda_4 = 1-\tan\psi \tag{7.2}$$

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 compared in the previous section (6.6) 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 7.1. *If $\mu > \sqrt{2}$ then for all but a countable set of $\theta \in (\hat{\theta}, \theta_{cc})$ all orbits of (5.13) are minimal.*

Putting Lemmas 6.2 and 7.1 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 anologue of the dichotomy of rigid rotations between irrational rotations (all orbits dense) and rational rotations (all orbits periodic).

8. 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 $x \to x+\alpha\ (mod\ 1)$ 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

$$\tan\theta + \tan\psi = \tfrac{1}{4}, \quad \mu > \sqrt{2} \tag{8.1}$$

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 $\mu > \sqrt{2}$ (8.1) corresponds to case A of previous sections and the induced map is H defined in (5.13). Putting (8.1) and (5.13) together shows that $H(x) = x - \frac{1}{4}$ if $x \in [\frac{1}{4}, 1) = I_4$ and then (5.9), (5.10), (5.13) and (8.1) 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 (6.5).

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

$$\alpha^4 - 2k\alpha^3 + (1+k^2)\alpha^2 - \tfrac{2k\mu^2}{\mu^2-1}\alpha + k^2\tfrac{\mu^2}{\mu^2-1} = 0 \tag{8.2}$$

with $k = \frac{1}{4}$. Equations (5.13) and (8.1) 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 $[\tan\psi, \frac{1}{4})$

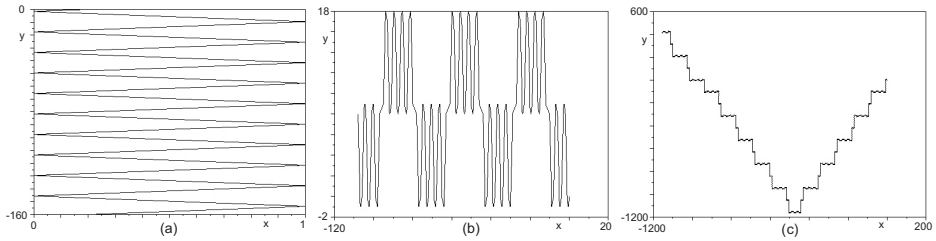

FIG. 6. *Light rays with* $\mu = 1.5$. *(a)* $x_0 = 0.2$*;* $\theta_0 = 0.0992981$*; 100 incidence events; (b)* $x_0 = 0.1$*;* $\theta_0 = 0.0992981$*; 300 incidence events;* $x_0 = 0.1$*;* $\theta_0 = 0.1$*; 5000 incidence events.*

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 (8.1) and $\mu = 1.5$ (which is the refractive index of some glass). For this value numerical solution of (8.2) gives

$$\begin{aligned} &\alpha = \tan\psi \approx 0.1503743, \quad && \psi \approx 0.1492560, \\ &\theta \approx 0.0992981, && \tan\tilde{\theta} \approx 0.8767620. \end{aligned} \tag{8.3}$$

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 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 $|\theta| \to \frac{\pi}{2} - |\theta|$, 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.

9. 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 refration 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 arrangements 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] C.P. Dettmann and O. Georgiou (2012) Quantifying intermittency in the open drivebelt billiard *Chaos* **22** 02611, doi:10.1063/1.3685522

[6] M. Keane (1975) Interval exchange transformations *Math. Z.* **141** 25-31.

[7] A. Lopes and R. Markarian (1996) Open Billiards: Invariant and Conditionally Iinvariant Probabilities on Cantor Sets *SIAM J. Appl. Math.* **56** 651–680.

[8] 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.

[9] M. Rees (1981) An alternative approach to the ergodic theory of measured foliations on surfaces *Ergod. Th. Dyn. Sys.* **1** 461-488.

[10] 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. Sina, 360–382.

[11] S. Tabachnikov (2005) *Geometry and Billiards*, Student Mathematical Library 30, AMS, Providence, RI.

[12] W. A. Veech (1978) Interval exchange transformations, *J. Analyse Math.* **33** 222-272.

[13] W. A. Veech (1986) The Teichmüller geodesic flow *Ann. Math.* **124** , 441-530.

[14] M. Viana (2006) Ergodic Theory of Interval Exchange Maps *Rev. Mat. Complut.* **19** 7–100.

[15] J.C. Yoccoz (2010) Interval exchange maps and translation surfaces, *Clay Math. Proc.* **10** 1–70.