Ergod. Th. & Dynam. Sys. (2004), **24**, 89–106
DOI: 10.1017/S0143385703000269
© *2004 Cambridge University Press*
Printed in the United Kingdom

Finite horizon Riemann structures and ergodicity

VICTOR J. DONNAY† and CHARLES PUGH‡
† *Mathematics Department, Bryn Mawr, PA 19010, USA*
(e-mail: vdonnay@brynmawr.edu)
‡ *Mathematics Department, University of California, Berkeley, CA 94720, USA*
(e-mail: pugh@math.berkeley.edu)

(*Received* 13 *February* 2003 *and accepted in revised form* 20 *May* 2003)

Abstract. In this paper we show that any surface in $\mathbb{R}^3$ can be modified by gluing on small 'focusing caps' so that its geodesic flow becomes ergodic. A new concept, finite horizon cap geometry, is what makes the construction work.

1. *Introduction*

In [**2**], it was shown that any surface embeds in $\mathbb{R}^3$ in such a way that the geodesic flow of its induced Riemann structure is ergodic. The construction involves gluing focusing caps to a surface of non-positive curvature. Here we use a new approach—the focusing caps are arranged to produce a *finite horizon* and thereby we eliminate the curvature assumption.

THEOREM A. *Given any surface $M \subset \mathbb{R}^3$, small disjoint focusing caps can be glued to M so that its geodesic flow becomes ergodic, has almost everywhere positive Lyapunov exponent and is Bernoulli.*

Throughout this paper, when we refer to a surface in $\mathbb{R}^3$ we mean a smooth, compact, boundaryless surface that is a subset of $\mathbb{R}^3$. We give it the Riemann structure it inherits from $\mathbb{R}^3$.

We say that a surface with caps as shown in Figure 1 has finite horizon geometry if every geodesic that begins off the caps eventually enters a cap. In the context of focusing caps, the finite horizon concept was first introduced in [**7**]. Previously, the notion was used in studying billiards on the plane that contained periodic arrays of scatterers [**1**]. In a non-mathematical context, it appears that the Romans used a similar notion of finite horizon when deploying their troops for battle: see Figure 2 (from [**13**, p. 14]).

Our strategy is as follows. We triangulate the surface M, subdivide it by the Freudenthal method [**10**], and glue caps to the resulting nearly flat and not too eccentric triangles using the finite horizon pattern shown in Figure 3. (Gluing is easier to analyze when the triangles

https://doi.org/10.1017/S0143385703000269 Published online by Cambridge University Press

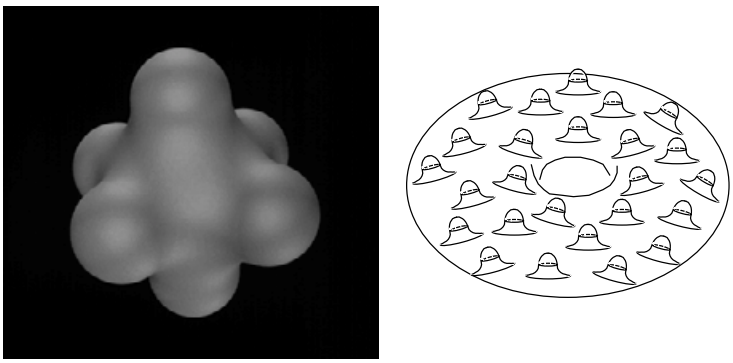


FIGURE 1. Comparison of the previous and the current capping possibilities.

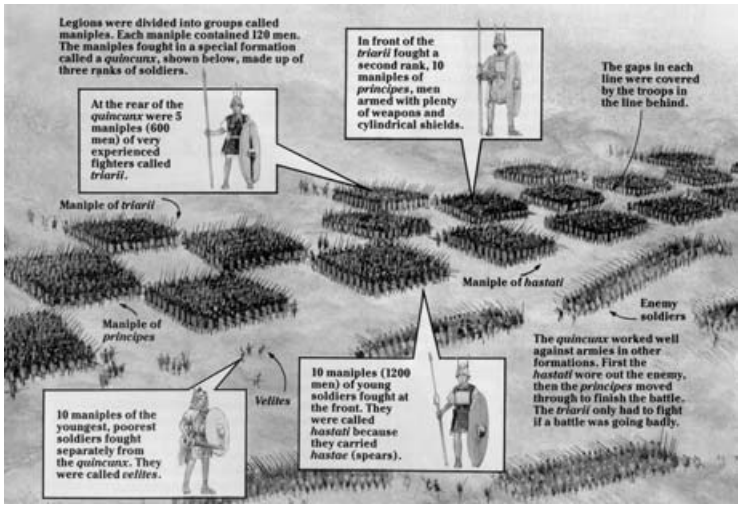


FIGURE 2. Finite horizon deployment of Roman troops. (Figure 2 is reproduced from 'The Romans' by permission of Usborne Publishing, 83–85 Saffron Hill, London EC1N 8RT. Copyright © 1990 Usborne Publishing Ltd.)

are of roughly unit size, so we scale the manifold accordingly. Scaling has no effect on ergodicity.) Then we apply the Burns–Gerber [**3**] adaptation of Wojtkowski's cone-field method [**17**] to prove the positive Lyapunov exponents [**15**] and the ergodicity and apply a result of Katok and Burns [**11**] to prove that the system is Bernoulli. We obtain the following result.

THEOREM B. *If M has a finite horizon geometry of focusing caps that satisfies a strictly invariant cone condition, then its geodesic flow is ergodic, has almost everywhere positive Lyapunov exponent and is Bernoulli.*

See §5 for the full statement and proof of Theorem B.

Using the notion of finite horizon geometry, we have created embedded surfaces of high genus for which the geodesic flow is Anosov (uniformly hyperbolic) [**9**]. In Appendix B we

https://doi.org/10.1017/S0143385703000269 Published online by Cambridge University Press

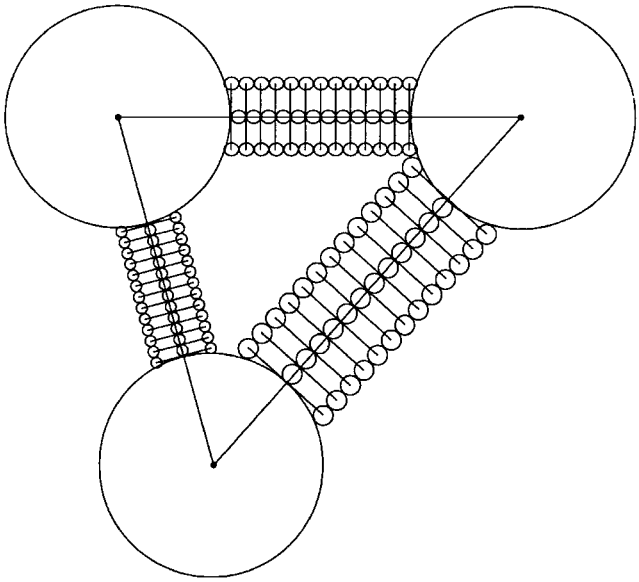


FIGURE 3. An array of base discs for focusing caps, which will be glued to the discs following this pattern. The discs are exaggerated in size; the number N of pearl discs is much larger than the twelve shown.

describe an extension of the main finite horizon construction that allows us to strengthen the result in [**9**] and to prove the following.

THEOREM C. *For all sufficiently large g, there exist embedded surfaces in R^3 of genus g for which the geodesic flow is Anosov.*

As a complement to our general construction, in Appendix A we produce some explicit examples of ergodic, finite horizon surfaces based on the flat torus, establishing a result announced in [**7**].

2. *Regular triangulation*

In this section, we construct a sequence of progressively finer smooth triangulations of M, which, in contrast to the barycentric sequence, are composed of triangles whose geometry is uniformly non-degenerate. The idea is due to Freudenthal [**10**].

Let T be an affine triangle. We always assume that T has diameter less than one. Its *first Freudenthal subdivision* consists of the four half-sized copies of T shown in Figure 4. Their vertices are the vertices of T and the edge midpoints. Iteration gives the nth Freudenthal subdivision of T, which divides it into 4^n subtriangles, each a copy of T reduced by the factor $1/2^n$.

Fix a smooth triangulation Φ of M. Thus Φ is a homeomorphism from a two-dimensional polytope P to M [**16**, p. 124]. The polytope consists of affine triangles T that meet only along common edges and at common vertices. Smoothness means that for each T, the restriction of Φ to T is a diffeomorphism from T to its image in M. That is, $\Phi|_T$ extends to a neighborhood U of T in the affine plane containing T, say $\Psi : U \to M$,

https://doi.org/10.1017/S0143385703000269 Published online by Cambridge University Press

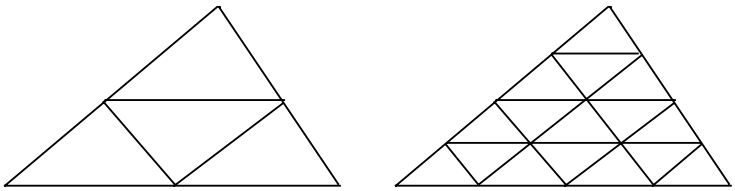


FIGURE 4. The first and second Freudenthal subdivisions of T.

and Ψ is a diffeomorphism from U to a neighborhood of T in M. Successive subdivision of each T in P gives polytopes P_n and *Freudenthal triangulations* $\Phi : P_n \to M$.

It is convenient to scale M by setting

$$M_n = \{2^n p : p \in M\}.$$

Then $\Phi_n = 2^n \cdot \Phi : P_n \to M_n$ triangulates M_n.

Let $\mathcal{T}_n$ denote the set of nonlinear triangles $\Phi_n(T_n)$ where T_n is an nth-order Freudenthal subtriangle of T, and T is a triangle of the original polytope P. The triangles in $\mathcal{T}_n$ have roughly unit size.

Let $\mathcal{L}$ be the set of all affine triangles $\pm(D\Phi)_z(T - z)$ where $z \in T$ and T is a triangle in P. The plus-or-minus sign is there to take care of the 'upside down' Freudenthal triangles. The notation $T - z$ indicates a translation of T to the origin. The natural parameterization of a triangle in $\mathcal{L}$ is

$$\lambda : u \mapsto (D\Phi)_z(u - z)$$

as u varies in T. The smoothness of $\Phi|_T$ and compactness of M imply that $\mathcal{L}$ is a compact set of triangles.

THEOREM 2.1. *As $n \to \infty$, $\mathcal{T}_n$ smoothly subconverges to $\mathcal{L}$.*

Proof. Smooth subconvergence means that for each $\epsilon > 0$ and each differentiability degree k, there is an n_0 such that if $n \geq n_0$ and τ_n is a natural parameterization of a suitable translate of $\Phi_n(T_n) \in \mathcal{T}_n$, then there is some $\lambda \in \mathcal{L}$ with

$$\|\tau_n - \lambda\|_{C^k} < \epsilon. \tag{1}$$

Assertion (1) amounts to Taylor's theorem in the form that states

$$\frac{f(z + tu) - f(z)}{t} \to (Df)_z(u)$$

as $t \to 0$.

Parameterize T_n by $R_n : T \to T_n$, where R_n shrinks T conformally to T_n, rotating by 180° if T_n is upside down. Let $z \in T_n$ be the fixed point of R_n. Then

$$R_n(u) = z \pm \frac{u - z}{2^n}.$$

As above, let $\lambda = \pm(D\Phi)_z(u - z)$ where the sign is determined by whether R_n is a dilation or a dilation and rotation. Then

$$\tau_n : u \mapsto \Phi_n \circ R_n(u) - \Phi_n(z)$$

https://doi.org/10.1017/S0143385703000269 Published online by Cambridge University Press

parameterizes a translate of the nonlinear triangle $\Phi_n(T_n)$ and

$$\begin{aligned}\tau_n(u) - \lambda(u) &= \Phi_n(R_n(u)) - \Phi_n(z) - (\pm(D\Phi)_z(u-z)) \\ &= \frac{\Phi(z \pm 2^{-n}(u-z)) - \Phi(z)}{2^{-n}} - (D\Phi)_z(\pm(u-z)).\end{aligned}$$

As $n \to \infty$ and u varies in T, Taylor's theorem gives uniform convergence to zero, verifying (1) with differentiability degree $k = 0$.

The derivative of $\tau_n - \lambda$ with respect to u is

$$(D\tau_n)_u - (\pm(D\Phi)_z)) = \pm((D\Phi)_{u_n} - (D\Phi)_z)$$

where $u_n = z \pm 2^{-n}(u - z)$. The continuity of $D\Phi$ implies that this tends uniformly to zero as $n \to \infty$, verifying (1) when the differentiability degree is $k = 1$. When $k \geq 2$, the kth derivative of $\tau_n - \lambda$ is

$$(D^k\tau_n)_u = 2^{-n(k-1)}(D^k\Phi)_{u_n}$$

which tends uniformly to zero as $n \to \infty$, completing the proof. □

3. *Finite horizon combinatorics*

A Riemann structure on M has *ϕ-finite horizon* with respect to a collection $\mathcal{C}$ of curves $C_1, \ldots, C_k$, if every unit geodesic meets at least one curve in $\mathcal{C}$ at an angle $\geq \phi$. In this section we construct such 'blocking curves'.

As a model in $\mathbb{R}^2$ we begin with the affine triangle T and the collection $\mathcal{C}$ of circles $C_1, \ldots, C_k$ that bound the discs $D_1, \ldots, D_k$ shown in Figure 3. We call the discs centered at the vertices of T *vertex discs*, the small discs along its edges *pearl discs* and the other discs *wing discs*. The vertex discs have radius R. There are N pearl discs along each edge of T and $2N + 2$ wing discs, all of radius $r \ll R$. (The radius r of the pearl discs along an edge is proportional to the length of the segment omitted by the vertex discs, so r may vary from edge to edge.) A vertex disc is tangent to two pearl discs and four wing discs. The wing discs come in pairs: the segments between paired wing disc centers are tangent to vertex discs and pearl discs and have length $2\sqrt{2rR + r^2}$. All the discs have disjoint interiors.

We use the following notation. Given a circle C with center p and radius r, its μ-dilation is the circle with center p and radius μr and is denoted by $C(\mu)$. Likewise, if $\mathcal{C} = \{C_1, \ldots, C_k\}$ we write

$$\mathcal{C}(\mu) = \{C_1(\mu), \ldots, C_k(\mu)\}.$$

PROPOSITION 3.1. *There exist constants $0 < \phi < \mu < 1$ such that the Euclidean Riemann structure on T has the ϕ-finite horizon property with respect to $\mathcal{C}(\mu)$, i.e. each unit segment σ with an endpoint in T meets at least one circle in $\mathcal{C}(\mu)$ at an angle $\geq \phi$.*

Proof 1. Inspect Figure 3. □

Proof 2. Suppose not. There are sequences ϕ_n, μ_n and σ_n such that $\phi_n \to 0$, $\mu_n \to 1$ and σ_n is a unit segment starting in T which only meets the circles in $\mathcal{C}(\mu_n)$ at angles less

https://doi.org/10.1017/S0143385703000269 Published online by Cambridge University Press

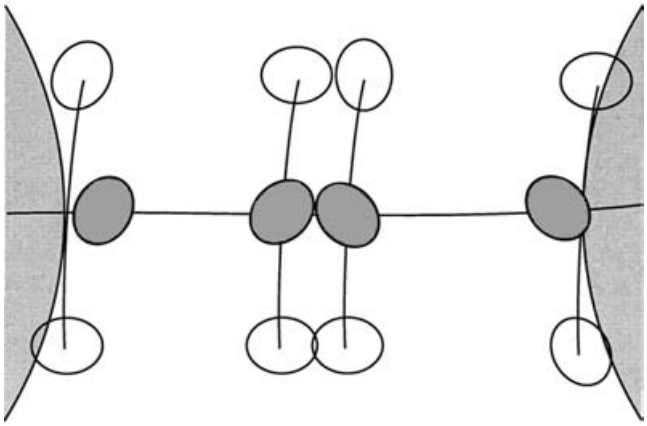


FIGURE 5. Nonlinear, g_n-geodesic blocking circles. The vertex discs are shaded lightly, the pearl discs heavily, and the wing discs are unshaded. All the discs are nearly round.

than or equal to ϕ_n. The compactness gives a subsequence of σ_n that converges to a unit segment starting in T and which crosses no circle in $\mathcal{C}$ at a positive angle.

Since σ has unit length, it crosses the boundary of T. If it does so inside a vertex circle or a pearl circle, it crosses that circle at a positive angle—a contradiction. Otherwise, σ crosses the boundary of T at a point of tangency between two pearl circles, or between a pearl circle and a vertex circle, and σ is tangent to both. However, then it crosses a wing circle orthogonally—a contradiction. □

As in §2 we set

$$M_n = \{2^n p : p \in M\}$$

and let $\Phi_n : P_n \to M_n$ be the nth Freudenthal triangulation of M_n, $\mathcal{T}_n$ its set of triangles, and g_n the Riemann structure that M_n inherits from $\mathbb{R}^3$. To construct a global blocking family $\mathcal{C}_n$ we proceed as follows; see Figure 5.

The vertex circles in $\mathcal{C}_n$. For each vertex $v_n \in M_n$ of the triangulation Φ_n, the vertex circle at v_n is the g_n-geodesic-circle $C(v_n, g_n, R)$ of radius R at v_n. Since the set of affine limit triangles $\mathcal{L}$ is compact, Theorem 2.1 implies that if $R \ll 1$ then (for each fixed n) the vertex circles $C(v_n, g_n, R)$ are disjoint and, if v_n is a vertex of $T_n \in \mathcal{T}_n$, then each edge of T_n contains an arc that joins one vertex circle to another.

The pearl circles in $\mathcal{C}_n$. Let I_n be the arc of an edge of a triangle $T_n \in \mathcal{T}_n$ that joins one vertex circle of $\mathcal{C}_n$ to another. Say I_n has endpoints a_n, b_n on the vertex circles and has length ℓ_n. On I_n we mark $2N + 1$ points that are equally spaced with respect to arclength

$$a_n = q_0, p_1, q_1, \ldots, p_N, q_N = b_n.$$

(The largeness of N is detailed below.) The arclength distance between successive marked points is

$$r = r(I_n) = \frac{\ell_n}{2N}.$$

The pearl circles along I_n in $\mathcal{C}_n$ are the g_n-geodesic circles with radius r and centers $p_1, \ldots, p_N$.

https://doi.org/10.1017/S0143385703000269 Published online by Cambridge University Press

The wing circles in $\mathcal{C}_n$. The wing circles along I_n also are g_n-geodesic circles of radius r. Their centers are the endpoints of g_n-geodesics of length $\sqrt{2rR + r^2}$ that leave I_n normally at $q_0, \ldots, q_N$.

If N is large then $r \ll R$, which makes the wing circles very close to the arcs I_n. By compactness of $\mathcal{L}$ and Theorem 2.1 we choose N large enough so that the wing circles along one edge are disjoint from those along the others.

Definition. $\mathcal{C}_n$ is the *Freudenthal blocking family* for M_n and $\mathcal{C}_n(\mu)$ is its μ-dilation.

Remark 3.2. Each affine triangle T in the compact set $\mathcal{L}$ supports a Freudenthal blocking family with the fixed values of R and N chosen above. There exist ϕ, μ such that the results of Proposition 3.1 hold uniformly for all $T \in \mathcal{L}$. For this value of μ, the circles in $\mathcal{C}_n(\mu)$ on M_n are disjoint for all sufficiently large n. As $n \to \infty$, the nonlinear configuration $\mathcal{C}_n$ becomes more and more linear. As space curves, the g_n-geodesic circles approach flat Euclidean circles. The constants R, N stay fixed. The disjointness of the circles in $\mathcal{C}(\mu)$ for all $L \in \mathcal{L}$ implies disjointness of the circles in $\mathcal{C}_n(\mu)$, when n is large.

4. *Caps*

A *focusing cap* is a surface of revolution S whose geodesics have a special symmetry. Expressed in cylindrical coordinates,

$$S = \{(r, \theta, z) : z = f(r) \text{ and } |r| \leq 1\}$$

where $f : \mathbb{R} \to [0, \infty)$ satisfies:

- $f(r) = f(-r)$;
- f is positive on $(-1, 1)$ and zero on $\mathbb{R} \setminus (-1, 1)$;
- f is smooth except at $\pm 1/2$ where $f'(\pm 1/2) = \mp\infty$.

(See Figure 6.) The inherited Riemann structure on S is required to have negative curvature on the *flank* $\{(r, \theta, f(r)) : 1/2 < r < 1\}$ of S, positive curvature on its *top* $\{(r, \theta, f(r)) : r < 1/2\}$ and zero curvature on the closed *meridian* geodesic $\Gamma = \{(r, \theta, f(r)) : r = 1/2\}$. Furthermore, the focusing property holds: every infinitesimal family of trajectories (i.e. a Jacobi field) that is strictly divergent when it enters the top, focuses once while in the top and then is strictly divergent when it leaves the top. The hemisphere as the top of a C^1-focusing cap was first used by Osserman [**14**] (see [**2, 4, 5**] for the existence of smooth focusing caps).

Locally, a smooth manifold can be expressed as a graph over its tangent plane. Since $M \subset \mathbb{R}^3$, we have a smooth outward unit normal field N on M which makes the graph representation explicit.

The compactness of M lets us make an initial scaling so that for each $p \in M$ and each $u \in T_pM$ with $|u| \leq 1$, there is an $h(p, u) \in [-1, 1]$ such that

$$p + u + h(p, u)N_p \in M.$$

The *graph chart* at p is the diffeomorphism

$$H : u \mapsto p + u + h(p, u)N_p$$

that sends the unit disc in T_pM to a unit neighborhood of p in M.

https://doi.org/10.1017/S0143385703000269 Published online by Cambridge University Press

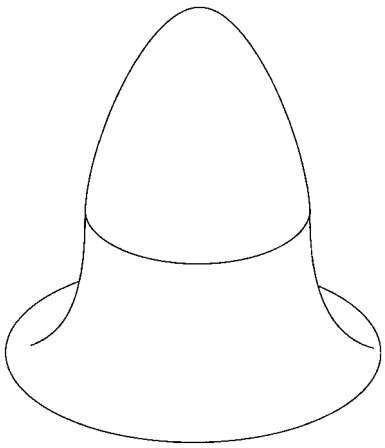


FIGURE 6. A focusing cap.

Next, we examine M under the scale doubling discussed in §2. We prefer unit sized caps on a large manifold to small caps on a unit sized manifold. As before we set

$$M_n = \{2^n p \in \mathbb{R}^3 : p \in M\}.$$

The unit graph chart for M_n at $2^n p$ is a scaled version of the graph chart at p, namely

$$H_n(p, u) = 2^n p + u + h_n(p, u) N_{2^n p},$$

where $u \in T_{2^n p} M_n$, $|u| \leq 1$ and $h_n(p, u) = h(p, u/2^n)$. Thus h_n and all its derivatives with respect to u tend uniformly to zero as $n \to \infty$. Here we are using the fact that $T_p M = T_{2^n p} M_n$ up to parallel translation. Uniformity refers to p varying in M. As a surface in $\mathbb{R}^3$, the image of the graph chart H_n becomes uniformly flat in the C^∞ sense as $n \to \infty$.

Now consider the triangulations Φ_n, their sets of triangles $\mathcal{T}_n$ and the Freudenthal blocking configurations $\mathcal{C}_n$ for M_n constructed in §§2 and 3. We assume that the initial triangulation Φ is chosen fine enough so that all triangles of Φ_n have diameter less than or equal to one.

Let $2^n p \in M_n$ be one of the centers for a blocking circle in $\mathcal{C}_n$. The circle has radius $\rho = r$ if it is a pearl or wing circle, and radius $\rho = R$ if it is a vertex circle. By construction and the initial scaling of M we have

$$0 < \inf \rho \leq \sup \rho < 1,$$

the infimum and supremum being taken over the circles in all the Freudenthal blocking families $\mathcal{C}_n$.

The difference between the graph chart neighborhood

$$M_n(p, \rho) = \{H_n(p, u) : |u| \leq \rho\}$$

of $2^n p$ in M_n and the disc bounded by the g_n-geodesic circle of radius ρ at $2^n p$ tends to zero as $n \to \infty$.

https://doi.org/10.1017/S0143385703000269 Published online by Cambridge University Press

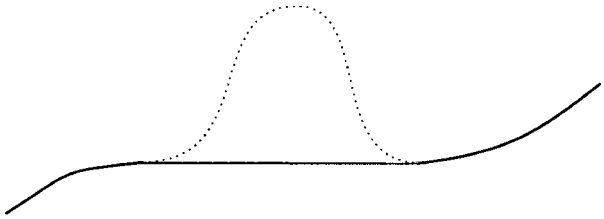


FIGURE 7. M_n with a plateau patch and cap, shown in cross section.

Let ϕ, μ be the constants of Proposition 3.1, referred to in Remark 3.2. Choose constants $0 < a < b < 1$ such that any straight line that crosses the (Euclidean) unit circle at an angle greater than or equal to ϕ crosses the circle of radius a at a positive angle. Then we fix a smooth bump function $\beta : \mathbb{R} \to [0, 1]$ such that $\beta(s) = 0$ when $|s| \leq a$ and $\beta(s) = 1$ when $s \geq b$.

Scale β down by the factor $\mu\rho$ and use it to construct a *plateau patch* at $2^n p \in M_n$,

$$\{2^n p + u + \beta(|u/\mu\rho|)h_n(p, u)N_{2^n p} : |u| \leq \rho\}.$$

(See Figure 7.)

The plateau patch contains a flat Euclidean disc of radius $a\mu\rho$ at $2^n p$. Plateau patches are disjoint because the discs bounded by the circles in $\mathcal{C}_n(\mu)$ are disjoint. Let

$$P_n : u \mapsto 2^n p + u + \beta(|u/\mu\rho|)h_n(p, u)N_{2^n p}$$

be the graph chart that represents the plateau patch. As $n \to \infty$, the plateau patch becomes uniformly C^∞-flat. The standard focusing cap is the graph $z = f(r)$ in cylindrical coordinates. We glue a scaled copy to a plateau patch as follows. Let

$$\widetilde{M}_n(p, \rho) = \{P_n(u) + f_\sigma(u)N_{2^n p} : |u| \leq \rho\},$$

where $\sigma = a\mu\rho$ and $f_\sigma(u) = \sigma f(|u|/\sigma)$. The qualitative properties of the focusing cap remain valid for the scaled copy. Let $\widetilde{M}_n$ be the surface M_n with the neighborhoods $M_n(p, \rho)$ replaced by the caps $\widetilde{M}_n(p, \rho)$ in this way. It is the capped surface whose geodesic flow we claim is Bernoulli.

THEOREM 4.1. *The family of cap boundary circles has the following form of the finite horizon property. There is a $\psi > 0$ such that if n is large and σ is a unit geodesic on $\widetilde{M}_n$ that starts in the uncapped part of $\widetilde{M}_n$ then it crosses a cap boundary circle at an angle greater than or equal to ψ, and it does so before reaching the meridian of a cap.*

Proof. Suppose not. Subsequences as in the proof of Proposition 3.1, but now also involving families of nonlinear triangles and Riemann structures on these triangles, lead to a smooth Riemann structure g defined on the complement of a family $\mathcal{C}(a\mu/2)$ of Euclidean discs and a unit g-geodesic γ that crosses no circle in $\mathcal{C}(a\mu)$ at a positive angle. The Riemann structure g is Euclidean outside the circles in $\mathcal{C}(a\mu)$ and γ starts outside them. Since γ crosses none of the circles at a positive angle, it is straight. Proposition 3.1 states that γ crosses a circle $C(\mu)$ in $\mathcal{C}(\mu)$ at an angle greater than or equal to ϕ. (See also Remark 3.2.) By the choice of a, it crosses the circle $C(a\mu)$ at a positive angle—a contradiction. □

https://doi.org/10.1017/S0143385703000269 Published online by Cambridge University Press

5. *Ergodicity*

In this section we prove Theorems A and B, the ergodicity of the geodesic flow φ for a capped surface.

The phase space of the geodesic flow for M is the unit tangent bundle SM. The bundle projection $\pi : SM \to M$ carries a φ-trajectory to its geodesic γ, the relation being $\varphi_t(u) = \gamma'(t)$.

Theorem 4.1 implies that every φ-trajectory except the closed meridian regularly enters the unit tangent bundle of the caps and thereby experiences fixed amounts of negative curvature. To be more precise, consider the unit cap in polar coordinates. Let ψ be the angle supplied by Theorem 4.1.

LEMMA 5.1. *In the unit cap's flank there is a flank annulus*

$$A = \{(r, \theta, f(r)) : r_1 < r < r_2\}$$

with $1/2 < r_1 < r_2 < 1$*, such that any unit geodesic entering the cap at an angle greater than or equal to* ψ *fully crosses* A *as it ascends in the cap. The geodesic takes no more than time* 1 *to completely pass through* A*.*

Proof. Choose r_1 very close to 1. □

Now consider the capped surface $\widetilde{M}$. Each cap contains a copy of the flank annulus A. Let $U \subset S\widetilde{M}$ be the set of unit vectors with basepoints that lie between the cap's flank annulus and its meridian. The set U is open and every φ-trajectory except the meridians meets it. That is,

$$\text{the } \varphi\text{-saturate of } U, \bigcup_{t\in\mathbb{R}} \varphi_t U, \text{ equals } S\widetilde{M} \text{ modulo a zero set.} \tag{2}$$

The Levi–Civita connection gives a Riemann structure on $S\widetilde{M}$ and an orthogonal splitting

$$T(S\widetilde{M}) = X \oplus N = X \oplus H \oplus V,$$

where X is the line field spanned by the geodesic spray (the vector field on $S\widetilde{M}$ that generates the geodesic flow), N is its field of normal planes, H is the horizontal subspace and V is the vertical subspace. The tangent to the geodesic flow, $T\varphi$, leaves the splitting $X \oplus N$ invariant. It does not leave the splitting $N = H \oplus V$ invariant (see [**12**, ch. 3.2]).

H and V are line fields. For $x \in SM$, let $P(x)$ be the standard, closed *positive cone* that consists of lines through the origin of $N(x)$ lying in the first and third quadrants with respect to $N = H \oplus V$.

THEOREM B. *Suppose that* M *has a finite horizon of focusing caps and the restriction of* P *to* U *is strongly invariant under* $T\varphi$ *in the following sense.*

- *(Weak invariance) If* $t > 0$ *and* $u, \varphi_t u \in U$ *then*

$$T\varphi_t P(u) \subset P(\varphi_t u).$$

- *(Eventual strict invariance) For each* $u \in U$ *there exists a positive time* $t = t(u)$ *such that*

$$T\varphi_t P(u) \subset \mathrm{Int}(P(\varphi_t u)).$$

Then φ *is ergodic, has almost everywhere positive Lyapunov exponent and is Bernoulli.*

https://doi.org/10.1017/S0143385703000269 Published online by Cambridge University Press

Proof. Under these hypotheses, the Burns–Gerber theorem [**3**] states that almost every point of U has a positive Lyapunov exponent and φ is locally ergodic on U in the sense that each connected component of U lies in a single ergodic component of φ plus a zero set.

The flow invariance of Lyapunov exponents and (2) imply that φ has almost everywhere positive Lyapunov exponent.

Consider any two caps C_1 and C_2, the corresponding connected components U_1 and U_2 of U, and the corresponding ergodic components E_1 and E_2. There is a geodesic γ from the top of C_1 to the top of C_2, which gives a φ-trajectory from U_1 to U_2. Perturbation gives a φ-trajectory from E_1 to E_2. The flow invariance of the ergodic components then implies that $E_1 = E_2$, so U is contained in a single ergodic component. By (2) φ is ergodic.

The geodesic flow for a surface is an example of a contact flow. Katok and Burns [**11**, Theorem 3.6] show that any ergodic contact flow with positive Lyapunov exponents is Bernoulli. Thus φ is Bernoulli. □

To check invariance of P, the following lemma is useful. It generalizes the two-dimensional Lobachevsky–Hadamard theorem stating that the positive cone field is uniformly strictly contracted everywhere for a surface of negative curvature.

LEMMA 5.2. *Let constants c, τ, C be given such that*

$$0 < c < \tfrac{1}{2}, \quad 0 < \tau < 1 < C.$$

Then there exists a constant $b > 0$ such that if $\gamma(t)$ is a unit speed geodesic with $t \in [0, \ell] \subset [0, 2]$ and the curvature $K(t)$ at $\gamma(t)$ satisfies:

- *for all $t \in [0, \ell]$, $-C \le K(t) \le b$;*
- *there exists an interval $[a, a + \tau] \subset [0, \ell]$ on which $K(t) \le -c$;*

then $T\varphi_\ell$ carries the closed positive cone at $\gamma'(0)$ into the open positive cone at $\gamma'(\ell)$. In particular, P is invariant when $K \le 0$.

Remark 5.3. To avoid trivial misunderstandings, the origin of a cone is always excluded from consideration.

Proof. We first take $b = 0$ and claim that $T\varphi_\ell$ carries P into its interior. As explained in [**12**], $T\varphi_t$ solves the equations

$$\dot{h} = v, \quad \dot{v} = -K(t)h, \tag{3}$$

where the coordinates (h, v) refer to $H \oplus V$. Let $H_+ = \{(h, v) : v = 0\}$ and $V_+ = \{(h, v) : h = 0\}$ be the horizontal and vertical edges of P. In terms of the slope $w = v/h$, (3) amounts to the Ricatti equation

$$\dot{w} = -K(t) - w^2.$$

Since $b = 0$, we have $\dot{w} \ge 0$ at $w_0 = 0$, which implies $w(a) \ge 0$. Over the time interval $[a, a + \tau]$ we have $\dot{w} \ge c - w^2$. If $w^2 \le c/2$, then $\dot{w} \ge c/2$, while if $c/2 < w^2 < c$, then $\dot{w} > 0$. Underestimating $w(a)$ by restarting at time $t = a$ with initial value $w(a) = 0$ gives

$$w(a + t) \ge \min(tc/2, \sqrt{c/2}).$$

https://doi.org/10.1017/S0143385703000269 Published online by Cambridge University Press

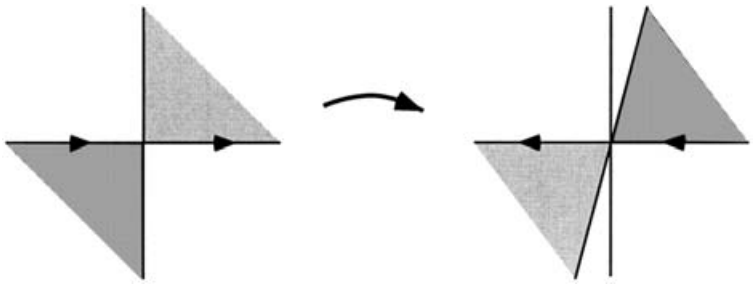


FIGURE 8. How passing over a cap's top rotates the positive cone into itself.

Since $\tau < 1$ and $c < 1/2$, we have that

$$w(a+\tau) \geq \frac{\tau c}{2}.$$

When $t \in [a+\tau, \ell]$ we have $\dot{w} \geq -w^2$ and $w(t)$ can decrease. The decrease to the underestimate for $w(a+\tau)$ is at worst $-2(\tau c/2)^2$, so

$$w(\ell, b=0) \geq \frac{\tau c}{2} - 2\left(\frac{\tau c}{2}\right)^2 = \frac{\tau c}{2}(1-\tau c) \geq \frac{\tau c}{4}.$$

This means that when $b = 0$ and $t > 0$, $T\varphi_t$ carries H_+ inside P and when $t = \ell$, this line has slope $\geq \tau c/4$.

V_+ is easier to analyze. Since $\dot{h} = v$ and $\dot{v} = -Kh = 0$ at $t = 0$, V_+ initially moves strictly inside P. Since $T\varphi_t(H_+) \subset P$ and $T\varphi_t$ preserves orientation, $T\varphi_t(V_+)$ remains interior to P. Hence $T\varphi_\ell(P) \subset \operatorname{int} P$ when $b = 0$.

By continuity the strict invariance of the cone remains true for small $b > 0$. □

Remark 5.4. Although the focusing cap has purely positive curvature, it still gives a form of this cone invariance. Namely, if the geodesic γ crosses the cap's top in time t, joining points on the meridian geodesic, then $T\varphi_t$ carries the positive cone at $\gamma'(0)$ into the positive cone at $\gamma'(t)$. It is able to do so because the geodesics focus once as they pass through the cap. The effect on the cone is that the first and third quadrants interchange and the horizontal reverses (see Figure 8).

THEOREM A. *When n is large, the geodesic flow for the capped surface $\widetilde{M} = M_n$ constructed in §4 is ergodic, has positive Lyapunov exponent almost everywhere and is Bernoulli.*

Proof. It suffices to check the hypothesis of Theorem B for the capped surface—strong invariance of the positive conefield P restricted to U. (Recall that $U \subset S\widetilde{M}$ and the basepoint of $u \in U$ lies in a cap's flank between the flank annulus and the meridian.)

Since $\widetilde{M}$ has negative curvature in the caps' flanks, *every* small positive time t has the property that $T\varphi_t$ carries $P(u)$ to the interior of $P(\varphi_t u)$ when $u \in U$. That is, the eventual strict invariance is automatic due to short time dynamics. It remains to check the weak invariance.

https://doi.org/10.1017/S0143385703000269 Published online by Cambridge University Press

Choose $c \in (0, 1/2)$ and $0 < \tau < 1 < C$ such that for all the capped surfaces M_n:

- on all the flank annuli A in all the caps, $K \leq -c$;
- the width of each flank annulus is greater than or equal to τ;
- at all points of M_n, $-C \leq K \leq C$.

Let b be the constant supplied by Lemma 5.2 and fix $\widetilde{M} = M_n$ with n so large that the curvature in the uncapped part of M_n is less than or equal to b.

Consider the forward φ-trajectory $\varphi_s(u)$ through $u \in U$, $0 \leq s \leq t$, and let $\gamma(s) = \pi(\varphi_s(u))$ be its geodesic.

While $\varphi_s(u)$ remains in U, the weak invariance of P under $T\varphi_t$ follows from the negative curvature of the flank. If $\varphi_s(u)$ leaves U by moving up the cap, but does not cross the meridian before returning to U, then γ stays in negative curvature and the weak invariance holds. If $\varphi_s(u)$ leaves and re-enters U because γ crosses the cap's top, then the weak invariance of P under $T\varphi_t$ follows from the cap's focusing property (see Figure 8). Finally, $\varphi_s(u)$ may leave and re-enter U because γ exits the cap at $\gamma(s_1)$, enters another cap at $\gamma(s_2)$, crosses that cap's flank annulus, enters U at $\gamma(s_3)$ and arrives at $\varphi_t(u) \in U$ without leaving U again. By construction

$$s_2 - s_1 \leq 1, \quad s_3 - s_2 \leq 1.$$

See Lemma 5.1 and recall that all the triangles have diameter less than or equal to 1. The flank's negative curvature gives the invariance of P from $\gamma(0)$ to $\gamma(s_1)$ and from $\gamma(s_3)$ to $\gamma(t)$, while Lemma 5.2 gives the invariance from $\gamma(s_1)$ to $\gamma(s_3)$. Thus P is invariant under the composite

$$T\varphi_t = T\varphi_{t-s_3} \circ T\varphi_{s_3-s_1} \circ T\varphi_{s_1}.$$

This completes the proof that P is weakly invariant with respect to $T\varphi$ and hence, by Theorem B, φ has the ergodic properties claimed in Theorem A. □

6. *Final remarks*

Here are some thoughts that occurred to us while writing this paper.

(a) Except for the construction of the cap itself, there is nothing intrinsically two dimensional about our proof. Nor is it necessary that the surface be embedded in $\mathbb{R}^3$, nor, in fact, that it be embedded at all. The construction of finite horizon blocking discs is valid on an abstract Riemann manifold. An unpublished work of Lohkamp, in which he aims to construct ergodic Riemann structures for manifolds of arbitrary dimension, proposes results along these lines.

(b) Although the capped and uncapped surfaces are close in a C^0 sense as point sets, the capped surface being a small section of the normal bundle of the uncapped surface and although their distance functions (their metrics in the sense of metric spaces) are close in the C^0 sense, their Riemann structures are not close.

It is easy to see that the Riemann structures cannot be C^1 close. If they were, then trajectories starting at the same point in the unit tangent bundle stay close to one another for a long time as they evolve under the two different geodesic flows. However, the surface with caps contains short closed geodesics that remain forever inside the cap, while the uncapped surface has no long geodesics that lie in sets of small diameter. If we perturb

https://doi.org/10.1017/S0143385703000269 Published online by Cambridge University Press

the cap to destroy the closed geodesic and hence improve the closeness of the Riemann structures, then the system can become non-ergodic due to the existence of a stable elliptic periodic orbit [**8**].

It remains to determine whether the Riemann structures are C^0 close.

(c) Our theorem shows that having an ergodic geodesic flow is a dense property for embedded surfaces with respect to the C^0 topology. It is easy to show that having a non-ergodic flow is also a dense property in this topology; just replace the focusing caps by 'lightbulbs' as in [**5**].

(d) Discussions with Keith Burns led us to conjecture that the finite horizon condition on a capped surface does not by itself imply ergodicity.

What is the boundary between ergodic and non-ergodic finite horizon surfaces?

The ergodic examples we constructed all have a finite horizon structure that is homogenous over the entire surface and in which geodesics enter a cap very quickly. These examples can be considered as perturbations of a flat situation. A more general theory of ergodic finite horizon surfaces, that would handle 'non-flat' finite horizon patterns, can be developed along the following lines.

Consider a new flank annulus A' with radii $r_1 < r_2 = 1$ where r_1 is very close to one. Then let $U' \subset S\widetilde{M}$ be the set of unit tangent vectors whose basepoints lie in a cap's new flank annulus and which are oriented downward with respect to the cap. Lemma 5.1 and Theorem B hold with U' in place of U, thereby producing a larger class of ergodic finite horizon surfaces. The proof of this more general ergodicity theorem gets a bit technical as we have to study the singularity lines generated by geodesics that become forward asymptotic to the meridians and hence do not return to U'.

(e) As constructed, capped surfaces are smooth but not analytic. Does there exist an analytic version of the capped torus? Burns, Donnay and Gerber [**2, 4**] have had some ideas about this.

(f) The capped cylinder has zero curvature off the caps (here we exclude the flanks from the caps and 'cap' means 'top of the cap'), which is consistent with the fact that a doubly capped surface diffeomorphic to the sphere always has points of zero curvature. It is unclear whether there exists a surface $M \subset \mathbb{R}^3$ such that:

- M is diffeomorphic to the 2-sphere;
- M contains $n \geq 3$ isometric copies of the flankless cap above;
- M has negative curvature at all points off the caps.

A reasonable conjecture is that if $n = 3$, then the answer is 'no' and, in fact, that the points of non-negative curvature off the caps of such a 'triod' form a set of dimension greater than or equal to 1. If $n \geq 4$ and we permit a finite number of points off the caps at which the curvature is zero, then the answer appears to be 'yes'.

(g) Given an initial finite horizon configuration of caps that is not ergodic, can we always add more caps to the uncapped part so that the resulting surface has an ergodic geodesic flow?

https://doi.org/10.1017/S0143385703000269 Published online by Cambridge University Press

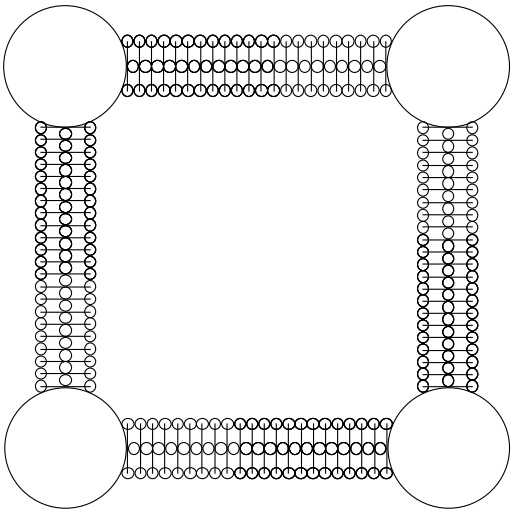


FIGURE A1. Finite horizon pattern on torus.

A. *Appendix. Explicit examples of ergodic, finite horizon surfaces generated using the flat torus*

Having developed a general theory of ergodicity for finite horizon surfaces, we now construct concrete examples. The simplest example comes from the flat torus.

There are many ways to give the flat torus a finite horizon geometry with blocking discs. See Figure A1, for example, where we use a rectangular pattern instead of a triangular one. We shrink the discs by a factor $\mu < 1$, making them disjoint, and glue on a cap to each.

THEOREM A.1. *The geodesic flow for the surface just constructed, the capped flat torus with finite horizon geometry, has positive Lyapunov exponent almost everywhere and is Bernoulli.*

Proof. Of course, this is a special case of the Main Theorem proved in §5, but the proof in this case is simpler. All the positive curvature is inside the caps, none on the uncapped part of the surface, so now it is *clear* that under the forward tangent geodesic flow, the positive cone field is preserved (sent into itself) while a geodesic remains outside the caps, is contracted into itself while the geodesic travels through a cap's flank and is preserved if the geodesic crosses a cap's top. Thus the Burns–Gerber theorem applies directly to the positive cone field and the issues of negative flank curvature outweighing positive uncapped curvature vanish from the proof. □

In [**5, 6**], it was shown that a flat torus with one or more caps, not necessarily satisfying the finite horizon condition, is Bernoulli. So Theorem A.1 is not new. However, if the flat capped torus is not finite horizon, then its geodesic flow becomes non-ergodic by a small perturbation outside the caps. For example, if the torus has only one cap that is not large enough to produce a finite horizon, then there are closed geodesics meeting no caps. They are parabolic, but become stably elliptic under a suitable perturbation of the Riemann structure, thus destroying the ergodicity.

https://doi.org/10.1017/S0143385703000269 Published online by Cambridge University Press

In contrast, the ergodicity for a flat, finite horizon capped torus is stable with respect to perturbations of the Riemann structure off the caps. This result was first announced in [**7**, §11]. The proof is now a one-liner: the finite horizon condition is open and the perturbed Riemann structure has at worst a negligible amount of positive curvature off the caps.

Using the rectangular capping pattern shown in Figure A1, it is also easy to visualize an embedding of the torus in 3-space that makes its geodesic flow ergodic. At first, the torus is embedded as a very large, locally nearly flat surface. The surface is covered by a periodic array of nearly flat rectangles, each of which is equipped with the finite horizon pattern of caps shown in Figure A1. As the embedded torus is scaled to be larger and more nearly flat, it continues to be covered by more copies of the same fundamental rectangle. The geodesic flow will be ergodic once the curvilinear rectangles are sufficiently close to flat. Since the embedded torus is covered by identical copies of one finite horizon rectangle, we avoid the compact family $\mathcal{L}$ of triangles (§2) which simplifies the analysis.

B. *Appendix. Embedded finite horizon surfaces with Anosov geodesic flow*

In [**9**], we used the techniques developed in this paper to prove the following.

THEOREM B.1. *There exist surfaces, of arbitrarily high genus, isometrically embedded in* $\mathbb{R}^3$ *for which the geodesic flow is Anosov.*

A modification of our finite horizon construction will allow us to prove the following stronger result.

THEOREM B.2. *There exists a constant N such that for every* $g \geq N$*, there exist surfaces of genus g that are isometrically embedded in* $\mathbb{R}^3$ *for which the geodesic flow is Anosov.*

To prove Theorem B.1, we took a sphere and triangulated it in the manner of §2 and marked the triangulation with a finite horizon pattern of disks as in §3. We took a second copy of this marked sphere, shrunk it slightly so that the two spheres were concentric and then connected adjacent pairs of disks by rotationally symmetric tubes of negative curvature, thereby producing a surface of high genus. We then followed the procedure of doubling the size of the spheres and refining the triangulations to produce a sequence of surfaces M_{g_n} in R^3 with monotonically increasing genus $g_n \to \infty$. Once the triangles in the triangulation become sufficiently uniformly flat, then the geodesic flow becomes Anosov. So given the sequence of surfaces M_{g_n}, there exists N such that for all $n \geq N$, the geodesic flow on the surface M_{g_n} is Anosov. These surfaces have a discrete set of genera $\{g_n\}, n \geq N$.

To prove Theorem B.2, given the discrete sequence of genera, we construct surfaces with genus g for all $g_n \leq g \leq g_{n+1}$ and show that there exists a $N > 0$ such that for all $g \geq N$, the geodesic flow on M_g is Anosov.

Recall (see Figure 3) our basic finite horizon construction on an affine triangle: we place vertex disks of radius R at each vertex, a string of N smaller disks (so called 'pearl' disks) along each edge between the vertex disks and $2N + 2$ 'wing discs' parallel to the string of pearl discs. These disks are tangential with their neighbors. Altogether this gives

https://doi.org/10.1017/S0143385703000269 Published online by Cambridge University Press

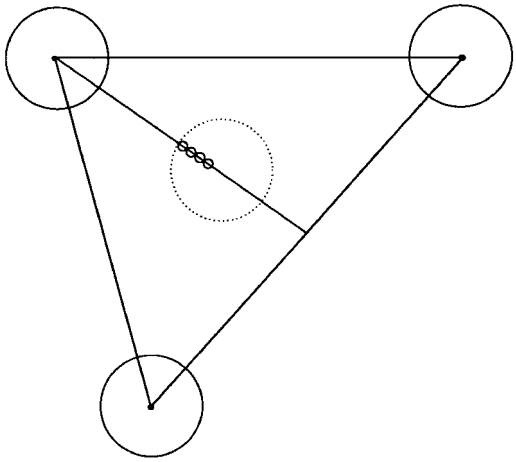


FIGURE B1. Adding k central disks to the finite horizon pattern.

$9(N+1)$ discs per triangle. After subdivision, each triangle will split into four smaller triangles which together have a total of $27(N+1)-3$ disks.

We modify our construction to have finite horizon patterns with k disks for each k between $9(N+1)$ and $27(N+1)-3$. We need to add these disks in some explicit way; the precise way is not particularly important. These families of disk patterns will allow us to produce surfaces of genus g for all values of g rather than just for a discrete set of values g_n.

We will place the additional disks in the center of the triangle. As a temporary guide to our construction, we place a fourth disk of radius R, based at the centroid of the triangle. Inside this disk, we place k smaller disks ('central disks') $0 \leq k \leq 18N+15$, each of radius $R/(18N+15)$. This choice of radius implies that $18N+15$ adjacent central disks will just fit inside the centroid disk. These central disks are placed side-by-side along the line joining a vertex to the midpoint of the opposite edge (see Figure B1). Once the central disks have been successfully inserted without intersecting any of the vertex or edge disks, we remove the temporary centroid disk.

We denote by $T_{(v_i,k)}, i = 1,2,3,\ 0 \leq k \leq 18N+15$ the affine triangle T equipped with the standard disk pattern along the edges together with k central disks placed along the line segment emanating from vertex v_i.

The affine triangles T we consider lie in a compact set $\mathcal{L}$ (§2) of the space of all affine triangles. We make the finite horizon construction in a uniform way over all these triangles. First we choose the radius R of the vertex disks and the centroid disk so that these four disks are disjoint for all $T \in \mathcal{L}$. We then determine how small the radius r of the edge disks must be so that the edge disks on adjacent sides are disjoint and so that the edge disks and centroid disk are disjoint. We then choose a value of N sufficiently large to ensure that the edge disks are sufficiently small.

Now, for this affine family of triangles $T_{(v_i,k)}$ with circles attached, we have a finite horizon result analogous to Proposition 3.1 and Remark 3.2.

https://doi.org/10.1017/S0143385703000269 Published online by Cambridge University Press

PROPOSITION B.3. *There exist constants* $0 < \phi < \mu < 1$ *such that the Euclidean Riemann structure on* $T_{(v_i,k)}$ *has the* ϕ*-finite horizon property with respect to* $\mathcal{C}(\mu)$. *The choice of* ϕ *and* μ *can be made uniform for all the affine triangles in* $T \in \mathcal{L}$, $i = 1, 2, 3$ *and* $0 \leq k \leq 18N + 15$.

The result follows by the same type of proof by contradiction using the compactness that was used in Proposition 3.1 and in Remark 3.2.

By using the finite horizon pattern with central disks, we can make finite horizon surfaces for all large genera. Once we have scaled the surface to be sufficiently large so that all the triangles in the triangulation are sufficiently flat, then the geodesic flow will be Anosov.

Acknowledgements. Charles Pugh thanks the Instituto de Matemática Pura e Aplicada for hospitality and support during part of the time this paper was written. Victor Donnay was a Visiting Scholar at the Department of Mathematics, University of California, Berkeley during 1998–99 when much of this paper was written. He thanks the department for their hospitality. He also thanks Keith Burns, Carlangelo Liverani and Maciej Wojtkowski for stimulating discussions and his son Michael for teaching him about the Roman legions.

REFERENCES

[1] L. A. Bunimovich and Ya. G. Sinai. Statistical properties of Lorentz gas with periodic configuration of scatterers. *Commun. Math. Phys.* **8** (1981), 479–497.
[2] K. Burns and V. J. Donnay. Embedded surfaces with ergodic geodesic flow. *Int. J. Bifurcation Chaos* **7**(7) (1997), 1509–1527.
[3] K. Burns and M. Gerber. Continuous cone families and ergodicity of flows in dimension three. *Ergod. Th. & Dynam. Sys.* **3** (1989), 19–25.
[4] K. Burns and M. Gerber. Real analytic Bernoulli geodesic flows on S^2. *Ergod. Th. & Dynam. Sys.* **9** (1989), 27–45.
[5] V. J. Donnay. Geodesic flow on the two-sphere, Part I: Positive measure entropy. *Ergod. Th. & Dynam. Sys.* **8** (1988), 531–553.
[6] V. J. Donnay. Geodesic flow on the two-sphere, Part II: Ergodicity. *Dynamical Systems*, Vol. 8 *(Lecture Notes in Mathematics, 1342)*. Springer, Berlin, 1988, pp. 112–153.
[7] V. J. Donnay. Physical examples of linked twist maps with chaotic dynamics. *Twist Mappings and Their Applications*. Eds. R. McGehee and K. Meyer. Springer, Berlin, 1993.
[8] V. J. Donnay. Elliptic islands in generalized Sinai billiards. *Ergod. Th. & Dynam. Sys.* **16** (1996), 975–1010.
[9] V. J. Donnay and C. Pugh. Anosov Geodesic Flows for Embedded Surfaces. *Astérisque (Geometric Methods in Dynamics (II))* **287** (2003), 61–69.
[10] H. Freudenthal. Simplizialzerlegungen von Beschrankter Flachheit. *Ann. Math.*, 2nd Series **43**(3) (1942), 580–582.
[11] A. Katok and K. Burns. Invariant cone families and stochastic properties of smooth dynamical systems. *Ergod. Th. & Dynam. Sys.* **14** (1994), 757–785.
[12] W. Klingenberg. *Riemannian Geometry*. de Gruyter, Berlin, 1982.
[13] A. Marks and G. Tingay. *The Romans*. The Usborne Illustrated World History, Usborne Publishing Ltd, England, 1990.
[14] R. Osserman. Lecture at the Geodesic Flow Workshop at Cal Tech, 1985.
[15] Ya. B. Pesin. Lyapunov characteristic exponents and smooth ergodic theory. *Russ. Math. Surveys* **32** (1977), 55–114.
[16] H. Whitney. *Geometric Integration Theory*. Princeton University Press, 1957.
[17] M. Wojtkowski. Invariant families of cones and Lyapunov exponents. *Ergod. Th. & Dynam. Sys.* **5** (1985), 145–161.

https://doi.org/10.1017/S0143385703000269 Published online by Cambridge University Press