

# CLOSED GEODESICS ON SURFACES WITHOUT CONJUGATE POINTS

VAUGHN CLIMENHAGA, GERHARD KNIEPER, AND KHADIM WAR

ABSTRACT. We obtain Margulis-type asymptotic estimates for the number of free homotopy classes of closed geodesics on certain manifolds without conjugate points. Our results cover all compact surfaces of genus at least 2 without conjugate points.

## 1. INTRODUCTION

1.1. **Main results.** Given a closed Riemannian manifold  $(M, g)$ , it is well known that each free homotopy class of loops contains at least one closed geodesic. Let  $P(t)$  denote the set of free homotopy classes containing a closed geodesic with length at most  $t$ , and  $\#P(t)$  its cardinality. From the point of view of dynamics and geometry it is of considerable interest to estimate  $\#P(t)$  as  $t$  tends to infinity.

**Theorem 1.1.** *Let  $(M, g)$  be a closed connected surface of genus at least 2 without conjugate points. Then*

$$(1.1) \quad \#P(t) \sim \frac{e^{ht}}{ht},$$

where  $h$  is the topological entropy of the geodesic flow on the unit tangent bundle  $SM$ , and the notation  $f_1 \sim f_2$  means  $\frac{f_1(t)}{f_2(t)} \rightarrow 1$  as  $t \rightarrow \infty$ .

In §1.2 we discuss the history of the Margulis asymptotics (1.1) and various related results, and in §1.3 we outline the proof, which uses ideas from the original work of Margulis [Mar69, Mar04] and from a recent preprint of Ricks [Ric19]. The following result clarifies the ingredients needed for our argument; precise definitions are in §2.

**Theorem 1.2.** *Let  $(M, g)$  be a closed Riemannian manifold without conjugate points that admits a background metric  $g_0$  of negative curvature and satisfies the divergence property (intersecting geodesics in the universal cover diverge). If the geodesic flow has a unique measure of maximal entropy and if this measure is almost expansive in the sense of Definition 2.8, then (1.1) holds.*

---

*Date:* August 4, 2020.

*2020 Mathematics Subject Classification.* 37C35, 37D40, 53C22.

V.C. is partially supported by NSF grant DMS-1554794.

G.K. and K.W. are partially supported by the German Research Foundation (DFG), CRC/TRR 191, *Symplectic structures in geometry, algebra and dynamics*.

We proved in [CKW19] that the hypotheses of Theorem 1.2 are satisfied for every surface of genus at least 2 without conjugate points, so this result implies Theorem 1.1.

In the course of the proof, we establish an equidistribution result for periodic orbits with lengths in  $(t - \epsilon, t]$ , which we believe to be of independent interest. The corresponding result for orbits with lengths in  $(0, t]$  was proved in [CKW19].

**Theorem 1.3.** *Let  $(M, g)$  be as in Theorem 1.2, and fix any  $\epsilon > 0$ . Given  $t > 0$ , let  $C(t)$  be any maximal set of pairwise non-free-homotopic closed geodesics with lengths in  $(t - \epsilon, t]$ , and consider the measure*

$$(1.2) \quad \nu_t = \frac{1}{\#C(t)} \sum_{c \in C(t)} \frac{\text{Leb}_c}{t},$$

where  $\text{Leb}_c$  is Lebesgue measure (length) along the curve  $c$  in the unit tangent bundle  $SM$ . Then as  $t \rightarrow \infty$ , the measures  $\nu_t$  converge in the weak\* topology to the measure of maximal entropy.

The results in [CKW19], and our arguments here, apply to a class of higher-dimensional manifolds as well; the following is a consequence of Theorems 1.2 and 1.3 together with [CKW19].

**Theorem 1.4.** *Let  $(M, g)$  be a closed Riemannian manifold without conjugate points that admits a background metric  $g_0$  of negative curvature and satisfies the divergence property. Suppose that the fundamental group  $\pi_1(M)$  is residually finite and that the entropy gap condition in §2.5 is satisfied. Then (1.1) holds, as does the equidistribution result in Theorem 1.3.*

**1.2. History.** Here we recall some of the history of (1.1) and weaker estimates, referring to the survey by Sharp [Mar04] for a more detailed overview and for a discussion of the stronger asymptotic estimates that are available in certain settings.

**1.2.1. Types of estimates.** We are interested in manifolds for which  $\#P(t)$  grows exponentially quickly. In this case one can identify three levels of asymptotic results, each stronger than the last.

- *Exponential growth rate:* existence of the limit

$$(1.3) \quad h = \lim_{t \rightarrow \infty} \frac{1}{t} \log \#P(t),$$

and identification of  $h$  with a dynamical or geometric quantity such as topological entropy or the growth rate of balls in the universal cover.

- *Uniform counting bounds:* existence of  $A, B > 0$  such that

$$(1.4) \quad \frac{A}{t} e^{ht} \leq \#P(t) \leq \frac{B}{t} e^{ht} \text{ for sufficiently large } t.$$

- *Multiplicative asymptotics:* the result in (1.1).

Broadly speaking, there are two types of approaches to proving these estimates. One approach relies on studying the dynamical properties of the geodesic flow on  $SM$ , with an important role played by the measure of maximal entropy. The other approach relies on methods from geometry, analysis, and geometric group theory, and leans heavily on the fundamental group  $\pi_1(M)$ , its isometric action on the universal cover of  $M$  as the group  $\Gamma$  of deck transformations, and the correspondence between free homotopy classes of curves on  $M$  and conjugacy classes in  $\Gamma \cong \pi_1(M)$ .<sup>1</sup> As will become clear later, our proof combines elements of both approaches.

1.2.2. *Negative curvature and uniform hyperbolicity.* When all sectional curvatures of  $(M, g)$  are negative, each free homotopy class is represented by a unique closed geodesic, so  $\#P(t)$  represents the number of closed geodesics with length at most  $t$ . In the 1950's Huber [Hub59, Satz 10] proved for closed surfaces of constant curvature  $-1$  that (1.1) holds with  $h = 1$ . He used a relation between the spectrum of the Laplacian and the length spectrum (the set of lengths of all closed geodesics) which in its general form is given by Selberg's trace formula [Sel56] and involves a sum over conjugacy classes in  $\Gamma$ . Proofs using Selberg's methods can be found in [McK72, Hej76, Bus92].<sup>2</sup> This formula extends to locally symmetric spaces of negative curvature in arbitrary dimension, where it was later used to deduce (1.1) by Gangolli [Gan77].

To deal with variable curvature, where there is no Selberg trace formula, different arguments are needed. In 1966, Sinai [Sin66] used dynamical methods to prove that if  $(M, g)$  has dimension  $n$  and all of its sectional curvatures lie between  $-b^2$  and  $-a^2$  for some  $b \geq a > 0$ , then the lower and upper limits in (1.3) lie between  $(n-1)a$  and  $(n-1)b$ ; however, he did not prove existence of the limit.

A striking improvement of Sinai's bounds was given by Margulis in his 1969 thesis, which established (1.1) – and hence (1.3) and (1.4) as well – for every  $(M, g)$  with (variable) negative sectional curvatures. Margulis's arguments are dynamical and apply to any topologically mixing Anosov flow, where in this dynamical setting  $P(t)$  denotes the set of periodic orbits with period at most  $t$ .<sup>3</sup>

At the heart of Margulis's proof is the construction of an invariant mixing measure  $m$  with the following properties:

---

<sup>1</sup>This correspondence means that one can obtain estimates on  $\#P(t)$  by counting conjugacy classes. It is worth pointing out that the question of growth of conjugacy classes is distinct from the question of growth of group elements; there are groups of exponential growth with only two conjugacy classes [Osi10, Corollary 1.2].

<sup>2</sup>In this context (1.1) is often called the prime number theorem for compact Riemann surfaces: note that the classical prime number theorem says that if  $\pi(x)$  is the number of primes less than or equal to  $x$ , then  $\pi(e^t) \sim \frac{e^t}{t}$ .

<sup>3</sup>This yields the geometric conclusion because the geodesic flow on  $SM$  is Anosov when  $M$  has negative curvature.

- (1) the conditional measures of  $m$  on stable and unstable manifolds contract and expand with a uniform rate under the flow;
- (2)  $m$  is a measure of maximal entropy for the flow;
- (3)  $m$  is the limiting distribution of periodic orbits in the sense of Theorem 1.3.

Margulis's results were announced without proof in [Mar69], in which the right-hand side of (1.1) had the form  $ct^{-1}e^{ht}$  for some unspecified constant  $c$ , which was stated to be equal to the entropy  $h$  in the constant curvature case. The stronger result that  $c = h$  in general was contained in his thesis itself, which was not published until 2004 [Mar04]. In the meantime, this equality was established in the 1984 thesis of Toll [Tol84], which gave an account of Margulis's ideas in English (albeit an unpublished one that only circulated informally); a description of Margulis's approach was also given in the 1995 book of Katok and Hasselblatt [KH95, §20.5–6].

After Margulis's work was announced but before his arguments were widely known, several important alternate dynamical approaches were developed. First came the work of Bowen in the early 1970's, which used the specification and expansivity properties satisfied by Anosov and Axiom A flows to prove (1.3) and (1.4) in [Bow72b] and [Bow72a, Theorem 4.1], respectively, and to show that periodic orbits are equidistributed with respect to an ergodic measure of maximal entropy. In [Bow74] Bowen proved that this measure of maximal entropy is in fact unique, and it is now referred to as the Bowen–Margulis measure. In 1983, Parry and Pollicott [PP83] gave an alternative proof of (1.1) based on symbolic dynamics and zeta functions, which covers Axiom A flows as well.

1.2.3. *Beyond negative curvature.* Outside of the uniformly hyperbolic setting, there are fewer comprehensive results available. In 1982 Katok used a closing lemma in Pesin theory to show that for a smooth flow with topological entropy  $h > 0$  and no fixed points, the exponential growth rate of the number of periodic orbits is at least  $h$  [Kat82, Theorem 4.1], yielding one inequality in (1.3). Recently, Lima and Sarig strengthened this to the lower uniform bound in (1.4) [LS19, Theorem 8.1] when the flow is on a manifold of dimension 3 (for geodesic flows, this requires  $M$  to be a surface so that  $\dim SM = 3$ ).

On the other hand, without the assumption of uniform hyperbolicity, it is easy to produce flows with uncountably many periodic orbits below a given length, which suggests that we should really be counting equivalence classes of periodic orbits, rather than individual orbits. In the most general setting it is not clear what equivalence relation would be appropriate, but for geodesic flow on  $SM$  it is natural to consider free homotopy of the underlying geodesics on  $M$ , and from now on we restrict our attention to this setting.

Now that  $\#P(t)$  counts free homotopy classes of closed curves, the upper bound in (1.3) holds without any further assumptions on  $M$  [Kni83, Satz 2.1]: growth of homotopy classes creates entropy. However, the lower bound in (1.3) no longer holds in general when we are counting homotopy classes instead of individual geodesics, as evidenced by the fact that the sphere admits metrics for which the geodesic flow has positive topological entropy [Don88, BG89]. Thus some assumptions must be placed on  $M$  to proceed further.

In 1983, the second author studied closed Riemannian manifolds with nonpositive curvature which admit a geodesic that does not bound a flat half plane [Kni83] (in particular, this includes manifolds of geometric rank 1) and proved the exponential growth rate (1.3), where  $h$  is the topological entropy. This used geometric methods, and in particular the result of Manning [Man79] that  $h$  gives the growth rate of the volume of balls in the universal cover.

In 1990, Kaimanovich [Kai90] showed that on a manifold with negative curvature, the Bowen–Margulis measure can also be obtained by building a family of measures on the boundary at infinity using a construction due to Patterson [Pat76] and Sullivan [Sul79]. The second author then adapted this to the rank 1 nonpositive curvature setting [Kni97, Kni02] and used it to obtain the uniform bounds (1.4).<sup>4</sup> In [Kni98], he showed that the measure of maximal entropy constructed from the Patterson–Sullivan measure is unique, and Babillot used the product structure of this measure to prove the mixing property [Bab02].

**Remark 1.5.** *An alternate proof of uniqueness in rank 1 based on a variant of Bowen’s specification property was recently given by the first author together with Burns, Fisher, and Thompson [BCFT18]. As with Bowen’s original work, this approach can be used to deduce the uniform estimates (1.4), although this is not done explicitly in the paper.*<sup>5</sup>

Beyond the Riemannian setting, Coornaert [Coo93] studied groups acting on Gromov hyperbolic metric spaces, and constructed Patterson–Sullivan measures on the Gromov boundary. Coornaert and the second author [CK02] used this together with ideas from [Kni83, Kni97] to prove that if a group  $\Gamma$  acts properly and cocompactly by isometries on a proper geodesic Gromov hyperbolic metric space whose Gromov boundary contains at least 3 points, then (1.4) holds for the set  $P(t)$  of conjugacy classes  $[\gamma]$  of elements  $\gamma \in \Gamma$  such that  $\inf_{x \in X} d(x, \gamma x) \leq t$ .<sup>6</sup>

---

<sup>4</sup>The upper bound in [Kni97] does not have the  $t$  in the denominator, but [Kni02, Proposition 5.3.6] explains how to obtain it.

<sup>5</sup>See [BCFT18, Corollary 4.8 and Proposition 6.4] for results toward (1.4).

<sup>6</sup>As with [Kni97], the upper bound in [CK02] has no  $t$  in the denominator, but it could be added with some extra work.

None of these results, however, establish the Margulis asymptotics (1.1) in any setting beyond negative curvature.<sup>7</sup> For surfaces that have negative curvature outside of a collection of radially symmetric ‘caps’, (1.1) was established by Weaver [Wea14], following Margulis’s original argument with suitable modifications. More recently, in an unpublished preprint [Ric19], Ricks has announced the result in the rank 1 CAT(0) setting, which includes the nonpositive curvature case; he follows Margulis’s approach, using geometric tools as in [Kni97, Kni98] to compensate for the lack of uniform hyperbolicity. As we describe below, we take a similar approach in this paper, although we make a number of substantial changes that allow us to sidestep locations where we were unable to follow the argument in [Ric19].

**1.3. Outline of the proof.** The proof of Theorem 1.2 uses ergodic theoretic properties of the measure of maximal entropy  $m$  for the geodesic flow  $\phi^t$  on  $SM$ . Beyond the basic geometric properties that we review in §2 (and prove in Appendix A), the key ingredients are as follows.

- (1) *Flow box and slice (§3)*: small sets  $B, S \subset SM$  with a product structure of (stable)  $\times$  (unstable)  $\times$  (flow), where  $B$  and  $S$  have the same cross-section in the first two components and have flow-lengths of  $\epsilon$  and  $\epsilon^2$ , respectively.<sup>8</sup>
- (2) *Closing lemma (§4)*: a nearly bijective correspondence between
  - the set of  $\epsilon$ -segments in which some  $c \in C(t)$  intersects  $B$ , where  $C(t)$  is as in Theorem 1.3, and
  - the set of components of  $S \cap \phi^{-t}B$ , indexed by a subset  $\Gamma(t)$  of the fundamental group.

In particular, the measures  $\nu_t$  from (1.2) have  $\nu_t(B) \approx \frac{\epsilon \#\Gamma(t)}{t \#C(t)}$ .

- (3) *Product structure and scaling properties (§§3.1 and 5.1)*: nearly every component of  $S \cap \phi^{-t}B$  has measure  $\approx e^{-ht}m(S)$ , and thus  $m(S \cap \phi^{-t}B) \approx \#\Gamma(t)e^{-ht}m(S)$ ;
- (4) *Mixing property for  $m$  (§5.2)*:  $m(S \cap \phi^{-t}B) \rightarrow m(S)m(B)$  as  $t \rightarrow \infty$ , so  $\#\Gamma(t) \sim e^{ht}m(B)$ .
- (5) *Equidistribution (§6.1)*:  $\nu_t \rightarrow m$  as in Theorem 1.3, giving the asymptotic  $\#C(t) \sim \frac{\epsilon \#\Gamma(t)}{t m(B)} \sim \epsilon \cdot e^{ht}/t$ .
- (6) *Integration (§6.2)*: Fixing  $b > 0$  and partitioning  $(b, T]$  into  $\epsilon$ -intervals with endpoints  $t_k$  gives a Riemann sum

$$\#P(T) \sim \#(P(T) \setminus P(b)) \sim \sum_k \epsilon \cdot \frac{e^{ht_k}}{t_k} \xrightarrow{\epsilon \rightarrow 0} \int_b^T \frac{1}{t} e^{ht} dt \sim \frac{1}{hT} e^{hT}.$$

<sup>7</sup>For rank 1 manifolds, the asymptotics were announced in the thesis of Gunesch [Gun02], but a complete proof was never published.

<sup>8</sup>Here we differ from prior approaches, which only used  $B$ .

These are the same basic ingredients as in Margulis’s original proof, with one notable exception: he works with components of  $B \cap \phi^{-t}B$ , which may have varying ‘depths’ between 0 and  $\epsilon$ , while we work with  $S \cap \phi^{-t}B$  and can disregard components that do not have full depth  $\epsilon^2$ . (This also makes our use the mixing property in §5 somewhat different from prior approaches.) It should also be pointed out that because we do not have uniform hyperbolicity, we follow [Kni97, Ric19] and work with the boundary at infinity instead of stable and unstable manifolds: this shows up in the definition of the flow box and slice, in the closing lemma, and in the product structure and scaling properties. Moreover, we carry out many of the counting steps in terms of the fundamental group  $\pi_1(M)$ .

## 2. DEFINITIONS AND GEOMETRIC PRELIMINARIES

**2.1. Closed curves.** Throughout the paper,  $(M, g)$  will be a closed Riemannian manifold. A (*smooth*) *closed curve* on  $M$  is a smooth map  $c: \mathbb{R}/\mathbb{Z} = S^1 \rightarrow M$ . Two such curves  $c_0, c_1: \mathbb{R}/\mathbb{Z} = S^1 \rightarrow M$  are *freely homotopic* if there exists a continuous map  $H: [0, 1] \times \mathbb{R}/\mathbb{Z} \rightarrow M$  such that  $H(0, t) = c_0(t)$  and  $H(1, t) = c_1(t)$  for all  $t$ . This is an equivalence relation on the set of closed curves, whose equivalence classes we call *free homotopy classes*.

By Cartan’s Theorem [dC92, Theorem 12.2.2], every free homotopy class of closed curves contains at least one closed geodesic, which minimizes length within the class. Thus (1.1) can be interpreted either as a result on the growth rate of free homotopy classes of closed geodesics, or of free homotopy classes of closed curves, since the two are equivalent. From now on we will work only with closed geodesics.

Since it is customary to parametrize geodesics with unit speed, we will write a closed geodesic as  $c: \mathbb{R}/\ell\mathbb{Z} \rightarrow M$ , where  $\ell > 0$  is the *length* of the closed geodesic. For manifolds without conjugate points (defined below), any two closed geodesics in the same free homotopy class have the same length; see [CKW19, Lemma 2.29] for the (short) proof.

Given a closed geodesic  $c$  and  $n \in \mathbb{N}$ , one can consider the closed geodesic  $c^n: \mathbb{R}/n\ell\mathbb{Z} \rightarrow M$  that consists of  $n$  orbits around  $c$ . A closed geodesic is *primitive* if it cannot be written as  $c^n$  for some closed geodesic  $c$  and  $n \geq 2$ . One can show that (1.1) is unaffected by whether we count all closed geodesics, or only primitive ones (see Remark 4.6).

As we will see in §2.4 below, there is a one-to-one correspondence between free homotopy classes of closed geodesics and conjugacy classes in the fundamental group. A conjugacy class is primitive if its elements cannot be written as proper powers. Note that it is possible to have a primitive closed geodesic whose free homotopy class corresponds to a non-primitive conjugacy class in the fundamental group: the boundary circle of a Möbius strip provides an example.

**2.2. Geodesic flows and standing assumptions.** Let  $SM$  be the unit tangent bundle of  $M$ , and  $\pi: SM \rightarrow M$  the canonical projection that takes each tangent vector to its basepoint. Each  $v \in SM$  determines a unique geodesic  $c_v: \mathbb{R} \rightarrow M$  with  $c_v(0) = \pi(v)$  and  $\dot{c}_v(0) = v$ . The *geodesic flow* on  $M$  is the flow  $\phi^t$  on  $SM$  given by

$$(2.1) \quad \phi^t(v) = \dot{c}_v(t).$$

Dynamical properties of the geodesic flow are related to curvature properties of the manifold, and in particular to the behavior of Jacobi fields, which govern how nearby geodesics evolve in time. Formally, a *Jacobi field* along a geodesic  $c: \mathbb{R} \rightarrow M$  is a vector field  $J(t)$  along  $c$  satisfying  $J'' + R(J, \dot{c})\dot{c} = 0$ , where  $R$  is the Riemann curvature tensor and  $'$  is covariant differentiation along  $c$ . Two points  $p, q$  on  $c$  are *conjugate* if there is a nonzero Jacobi field along  $c$  that vanishes at both  $p$  and  $q$ .

Throughout the paper, we consider manifolds without conjugate points. This class includes all manifolds with nonpositive curvature, as well as many manifolds with positively curved regions [Gul75]. It is also characterized by the property that the exponential map is nonsingular, or equivalently, that any two points in the universal cover are connected by a unique geodesic.

We impose two more standing assumptions:<sup>9</sup> we assume that  $M$  admits another Riemannian metric  $g_0$  for which all sectional curvatures are negative, and that  $M$  satisfies the divergence property in Definition 2.2 below. In the following sections we gather various definitions and results about such manifolds; some of the lemmas are quoted directly from the literature; the remainder are proved in Appendix A.

**2.3. Visibility manifolds and the boundary at infinity.** Write  $X$  for the universal cover of  $M$  and  $SX$  for the unit tangent bundle of  $X$  with respect to the lift of the Riemannian metric  $g$ . Let  $\text{pr}: X \rightarrow M$  be the canonical projection; we write  $\text{pr}_*: SX \rightarrow SM$  for the map this induces on the unit tangent bundles. From now on we will use an underline to denote objects in  $M$  and  $SM$ , so that for example  $\underline{c}$  will denote a geodesic on  $M$ , and  $c$  a geodesic on  $X$  that lifts  $\underline{c}$ , that is,  $\underline{c} = \text{pr} \circ c$ . We write  $\phi^t$  for the geodesic flow on both  $SM$  and  $SX$ .

Following [Ebe72], we can define the *boundary at infinity*  $\partial X$  provided the following condition is satisfied.

**Definition 2.1** ([Ebe72, Definition 1.3]). *A simply connected Riemannian manifold  $X$  without conjugate points is a uniform visibility manifold if for every  $\epsilon > 0$  there exists  $L > 0$  such that whenever a geodesic  $c: [a, b] \rightarrow X$  stays at distance at least  $L$  from some point  $p \in X$ ,*

---

<sup>9</sup>Note that these assumptions rule out some higher-dimensional manifolds with nonpositive curvature.

then the angle sustained by  $c$  at  $p$  is less than  $\epsilon$ , that is

$$\angle_p(c) = \sup_{a \leq s, t \leq b} \angle_p(c(s), c(t)) < \epsilon.$$

If  $M$  is a Riemannian manifold without conjugate points whose universal cover  $X$  is a uniform visibility manifold, then we say that  $M$  is a uniform visibility manifold.

The uniform visibility property implies the following *divergence property* [Ebe72, Proposition 1.5], and when  $M$  admits a negatively curved “background” metric  $g_0$ , the two conditions are equivalent [Ebe72, Theorem 5.1].<sup>10</sup>

**Definition 2.2.** *The manifold  $(M, g)$  has the divergence property if given any geodesics  $c_1 \neq c_2$  in the universal cover with  $c_1(0) = c_2(0)$ , we have  $\lim_{t \rightarrow \infty} d(c_1(t), c_2(t)) = \infty$ .*

For the remainder of the paper we assume that  $(M, g)$  has no conjugate points, admits a background metric of negative curvature, and satisfies the divergence property (and thus the uniform visibility property as well).

**Remark 2.3.** *In dimension 2, every surface of genus at least 2 without conjugate points admits a negatively curved metric, and the divergence property holds by [Gre56]. In higher dimensions the corresponding result is no longer true: there are rank 1 manifolds of nonpositive curvature, such as the graph manifolds described by Gromov [Kni98, §6], that admit totally geodesic isometric embeddings of  $\mathbb{R}^2$  and thus fail the uniform visibility condition.*

We describe a compactification of  $X$  following Eberlein [Ebe72].

**Definition 2.4.** *Two geodesic rays  $c_1, c_2: [0, \infty) \rightarrow X$  are called asymptotic if  $\sup_{t \geq 0} d(c_1(t), c_2(t)) < \infty$ . This is an equivalence relation; we write  $\partial X$  for the set of equivalence classes and call its elements points at infinity. We denote the equivalence class of a geodesic ray (or geodesic)  $c$  by  $c(\infty)$ .*

Given  $v \in SX$ , let  $c_v$  be the unique geodesic with  $\dot{c}_v(0) = v$ . The following is proved in [Ebe72, Propositions 1.5 and 1.14].

**Lemma 2.5.** *Given any  $p \in X$  and  $\xi \in \partial X$ , there is a unique geodesic ray  $c = c_{p, \xi}: [0, \infty) \rightarrow X$  with  $c(0) = p$  and  $c(\infty) = \xi$ . Equivalently, the map  $f_p: S_p X \rightarrow \partial X$  defined by  $f_p(v) = c_v(\infty)$  is a bijection. Moreover, for a given  $\xi$ , the initial vector  $\dot{c}_{p, \xi}(0)$  depends continuously on  $p$ .*

Following Eberlein we equip  $\partial X$  with a topology that makes it a compact metric space homeomorphic to  $S^{n-1}$ . Fix  $p \in X$  and let

---

<sup>10</sup>The divergence property is not mentioned in the statement of [Ebe72, Theorem 5.1], but the “Added in proof” section of that paper discusses its necessity.

$f_p: S_p X \rightarrow \partial X$  be the bijection  $v \mapsto c_v(\infty)$  from Lemma 2.5. The topology (sphere-topology) on  $\partial X$  is defined such that  $f_p$  becomes a homeomorphism. Since the map  $f_q^{-1} f_p: S_p X \rightarrow S_q X$  is a homeomorphism for all  $q \in X$  [Ebe72], the topology is independent of the reference point  $p$ .

The topologies on  $\partial X$  and  $X$  extend naturally to  $\bar{X} := X \cup \partial X$  by requiring that the map  $\varphi: B_1(p) = \{v \in T_p X : \|v\| \leq 1\} \rightarrow \bar{X}$  defined by

$$\varphi(v) = \begin{cases} \exp_p \left( \frac{v}{1-\|v\|} \right) & \|v\| < 1 \\ f_p(v) & \|v\| = 1 \end{cases}$$

is a homeomorphism. This topology, called the cone topology, was introduced by Eberlein and O’Neill [EO73] in the case of Hadamard manifolds and by Eberlein [Ebe72] in the case of visibility manifolds. In particular,  $\bar{X}$  is homeomorphic to a closed ball in  $\mathbb{R}^n$ . The relative topology on  $\partial X$  coincides with the sphere topology, and the relative topology on  $X$  coincides with the manifold topology.

**Remark 2.6.** *The manifold  $M$  is the quotient of  $X$  by the isometric action of the group  $\Gamma$  of deck transformations, which is isomorphic to the fundamental group  $\pi_1(M)$ . This extends to a continuous action on  $\partial X$ . The geodesic flow is topologically transitive [Ebe72], so every  $\Gamma$ -orbit in  $\partial X$  is dense, i.e., the action on  $X$  is minimal. Moreover, the Morse lemma (see for example [Kni02, Theorem 2.3]) gives a topological conjugacy between the actions of  $\Gamma$  on the boundaries at infinity with respect to the two metrics  $g$  and  $g_0$ ; in particular, every element  $\gamma \in \Gamma \setminus \text{Id}$  has exactly two fixed points on  $\partial X$ , since this is true for  $g_0$ .*

**Definition 2.7.** *Given  $\gamma \in \Gamma \setminus \text{Id}$ , let  $\xi_\gamma^- \in \partial X$  be the repelling fixed point of  $\gamma$ , and  $\xi_\gamma^+ \in \partial X$  be the attracting fixed point of  $\gamma$ .*

Every geodesic  $c: \mathbb{R} \rightarrow X$  determines two distinct points  $c(-\infty)$  and  $c(\infty)$  on  $\partial X$ . We will also use the notation

$$(2.2) \quad v^- := c_v(-\infty) \quad \text{and} \quad v^+ := c_v(\infty).$$

It is useful to write

$$(2.3) \quad \partial^2 X := \{(\xi, \eta) \in (\partial X)^2 : \xi \neq \eta\}$$

and to consider the *endpoint map*

$$(2.4) \quad E: SX \rightarrow \partial^2 X, \quad v \mapsto (v^-, v^+).$$

By [Ebe72, Proposition 1.7], every pair of distinct points on  $\partial X$  is joined by at least one geodesic, so the map  $E$  is onto. Note that  $E$  is continuous by the definition of the topology on  $\partial X$ .

It is sometimes important to work with pairs of points at infinity that are joined by *exactly* one geodesic.

**Definition 2.8.** *The expansive set of  $X$  is*

$$(2.5) \quad \mathcal{E} := \{v \in SX \mid \phi^{\mathbb{R}}(v) = E^{-1}(v^-, v^+)\}.$$

*We say that an invariant Borel probability measure  $\mu$  on  $SM$  is almost expansive if  $\mu(\text{pr}_* \mathcal{E}) = 1$ .*

**2.4. Periodic orbits and deck transformations.** There is a one-to-one correspondence between closed geodesics on  $M$  and conjugacy classes in the group  $\Gamma$  of deck transformations. Here we recall some of the elements of this correspondence that we will need.

Given a closed geodesic  $\underline{c}: \mathbb{R}/\ell\mathbb{Z} \rightarrow M$ , the length  $\ell > 0$  is such that  $\underline{c}(t + \ell) = \underline{c}(t)$  for all  $t \in \mathbb{R}$ , and thus for every lift  $c$  of  $\underline{c}$  there is a unique  $\gamma \in \Gamma$  such that

$$(2.6) \quad c(t + \ell) = \gamma c(t) \text{ for all } t \in \mathbb{R}.$$

**Definition 2.9.** *If  $c$ ,  $\gamma$ , and  $\ell > 0$  are such that (2.6) holds, then we say that  $\gamma$  is the axial isometry of  $(c, \ell)$ , and  $c$  is an axis of  $\gamma$ .*

If  $c$  is an axis of some  $\gamma$ , then it follows immediately that  $\underline{c} = \text{pr} \circ c$  is a closed geodesic on  $M$ . Moreover, every axial isometry  $\gamma$  of  $c$  (with any value of  $\ell$ ) fixes  $c(\pm\infty)$ , and by Remark 2.6 and Definition 2.7 this proves the following lemma.

**Lemma 2.10.** *If  $c$  is an axis of  $\gamma$ , then  $c(-\infty) = \xi_\gamma^-$  and  $c(\infty) = \xi_\gamma^+$ .*

**Remark 2.11.** *There may be multiple geodesics connecting  $\xi_\gamma^\pm$  besides  $c$  itself; these geodesics may or may not be axes of  $\gamma$ .*

It follows from Preissman's Theorem that in the background metric  $g_0$ , given every pair of points on  $\partial X$ , the set of  $\gamma$  that fixes these two points is either trivial or an infinite cyclic subgroup [dC92, Lemma 12.3.5]. Using the topological conjugacy from Remark 2.6 we get the same result for  $g$ .

**Lemma 2.12.** *Given any geodesic  $c$  on  $X$  such that  $\underline{c} = \text{pr} \circ c$  is closed, the set of  $\gamma \in \Gamma$  fixing  $c(-\infty)$  and  $c(\infty)$  is an infinite cyclic subgroup.*

**Remark 2.13.** *It is possible for the subgroup in Lemma 2.12 to contain non-identity elements for which neither  $c$  nor its reverse is an axis; for example, this occurs when  $c$  lifts a boundary circle of a Möbius strip.*

The preceding discussion lets us go from closed geodesics to deck transformations. We will also need to go in the other direction.

**Definition 2.14.** *Given  $\gamma \in \Gamma \setminus \text{Id}$ , the length of  $\gamma$  is*

$$|\gamma| := \inf\{d(q, \gamma q) : q \in X\}.$$

**Lemma 2.15.** *For every  $\gamma \in \Gamma \setminus \text{Id}$ , there exists  $q_0 \in X$  such that  $d(q_0, \gamma q_0) = |\gamma| \geq 2 \text{inj}(M)$ . Moreover, if  $c: \mathbb{R} \rightarrow X$  is a geodesic joining  $q_0$  and  $\gamma q_0$ , then  $c(t + |\gamma|) = \gamma c(t)$  for all  $t \in \mathbb{R}$ , so  $\underline{c} := \text{pr} \circ c$  is a closed geodesic.*

Lemma 2.15 is standard: the idea of the proof is in [BC64, §11.7, Theorem 10], and a complete proof is in [Sak96, pages 196–197]. We need two more standard results relating free homotopy classes, axial isometries, and endpoints on  $\partial X$ ; for completeness we provide proofs in Appendix A.

**Lemma 2.16.** *Let  $\underline{c}_0: \mathbb{R}/\ell_0\mathbb{Z} \rightarrow M$  and  $\underline{c}_1: \mathbb{R}/\ell_1\mathbb{Z} \rightarrow M$  be closed geodesics on  $M$  that lie in the same free homotopy class.<sup>11</sup> Then given any lift  $c_0$  of  $\underline{c}_0$ , there is a lift  $c_1$  of  $\underline{c}_1$  such that  $(c_0, \ell_0)$  and  $(c_1, \ell_1)$  have the same axial isometry, and hence the same endpoints.*

**Lemma 2.17.** *If  $\gamma$  is the axial isometry for both  $(c_0, \ell_0)$  and  $(c_1, \ell_1)$ , then  $\ell_0 = \ell_1 = |\gamma|$ , and the corresponding closed geodesics lie in the same free homotopy class.*

We remark that these lemmas (apart from the claim about endpoints on  $\partial X$ ) are true for general closed manifolds, although for convenience our proof of Lemma 2.17 uses  $\partial X$  via Lemma 2.10.

**2.5. Other geometric hypotheses.** In [CKW19] we proved existence and uniqueness of the measure of maximal entropy for a geodesic flow on a closed Riemannian manifold  $M$  without conjugate points under the following conditions.

- (1) There exists a Riemannian metric  $g_0$  on  $M$  for which all sectional curvatures are negative.
- (2) The visibility axiom is satisfied.<sup>12</sup>
- (3) The fundamental group  $\pi_1(M)$  is *residually finite*: the intersection of its finite index subgroups is trivial.
- (4) There exists  $h_0 < h$  such that any ergodic invariant Borel probability measure  $\mu$  on  $SM$  with entropy  $> h_0$  is almost expansive.

As discussed in [CKW19, Remarks 2.1 and 3.2], the first of these conditions is a genuine topological restriction, and excludes certain manifolds of nonpositive curvature that still obey the Margulis asymptotics by [Ric19]. It is not known whether or not the remaining three conditions can fail if the first holds. All four conditions hold for every surface of genus  $\geq 2$  [CKW19, §3].

In addition to proving existence and uniqueness of the MME, we proved in [CKW19] that it is mixing and has a product structure and scaling properties that we describe in the next section.

### 3. SCALING PROPERTIES AND PRODUCT STRUCTURE

#### 3.1. Busemann functions and conformal densities.

<sup>11</sup>Here we specify the periods  $\ell_0, \ell_1$  explicitly, as in §2.1; notice that replacing  $\ell_j$  with  $n\ell_j$  would change the free homotopy class.

<sup>12</sup>The result in [CKW19] is formulated using the divergence property, but they are equivalent when there is a negatively curved background metric.

**Lemma 3.1** ([Esc77, Proposition 1]). *Given  $v \in SX$  and  $q \in X$ , the limit*

$$b_v(q) := \lim_{t \rightarrow \infty} (d(q, c_v(t)) - t)$$

*exists and defines a  $C^1$  function on  $X$ . Moreover, we have*

$$\text{grad } b_v(q) = \lim_{t \rightarrow \infty} \text{grad} (d(q, c_v(t)) - t).$$

The function  $b_v$  is called the *Busemann function* associated to  $v$ . It is in fact  $C^{1,1}$  [Kni86, Satz 3.5].

**Definition 3.2.** *Given  $p \in X$  and  $\xi \in \partial X$ , let  $v \in S_p X$  be the unique unit tangent vector at  $p$  such that  $c_v(\infty) = \xi$ . We call  $b_\xi(q, p) := b_v(q)$  the Busemann function based at  $\xi$  and normalized by  $p$  ( $b_\xi(p, p) = 0$ ).*

The zero set of the Busemann function is the horosphere through  $p$  centered at  $\xi$ , say  $H_\xi(p)$ , and  $b_\xi(q, p)$  can be interpreted as the distance you need to travel along the geodesic from  $q$  towards  $\xi$  in order to reach this horosphere; see Figure 1. With this interpretation the following *cocycle property* of the Busemann function becomes transparent:

$$(3.1) \quad b_\xi(p, q) = b_\xi(p, r) + b_\xi(r, q) \text{ for all } \xi \in \partial X \text{ and } p, q, r \in X.$$

Note also that

$$(3.2) \quad b_\xi(p, q) = -b_\xi(q, p) \text{ for all } \xi \in \partial X \text{ and } p, q \in X.$$

Figure 1 also illustrates the function  $\beta_p: \partial^2 X \rightarrow (0, \infty)$  defined by

$$(3.3) \quad \beta_p(\xi, \eta) = -(b_\xi(q, p) + b_\eta(q, p)),$$

where  $q$  is a point on a geodesic  $c$  connecting  $\xi$  and  $\eta$ . Geometrically,  $\beta_p(\xi, \eta)$  is the length of the segment  $c$  which is cut out by the horoballs through  $(p, \xi)$  and  $(p, \eta)$ . Since  $\text{grad}_q b_\xi(q, p) = -\text{grad}_q b_\eta(q, p)$  for all points  $q$  on geodesics connecting  $\xi$  and  $\eta$ , this number is independent of the choice of  $q$ .

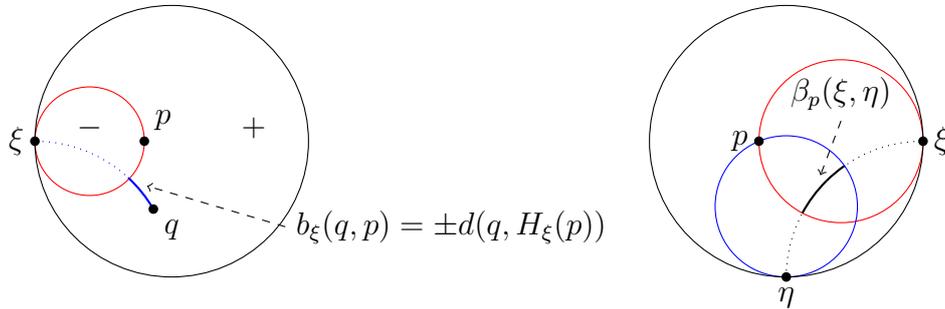


FIGURE 1. Geometric interpretation of  $b_p(q, \xi)$  and  $\beta_p(\xi, \eta)$ .

It is easy to see that  $b_\xi(p, q)$  is 1-Lipschitz in both  $p$  and  $q$  by either using the definition directly or by using (3.1) and the fact that  $|b_\xi(p, q)| \leq d(p, q)$ . We will also need to know how  $b_\xi(p, q)$  varies with

$\xi$ . This is equivalent to understanding how  $H_\xi(p)$  varies with  $\xi$ , which is accomplished by the following result; see [CKW19, Proposition 6.2] for the statement we use and for precise references to the proof, which can be found in [Pes77] and uses results from [Ebe72] (see also [Rug07, Lemma 4.11]).

**Proposition 3.3.** *Let  $(M, g)$  be a closed Riemannian manifold without conjugate points that satisfies the uniform visibility condition and admits a background metric of negative curvature, and let  $X$  be its universal cover. Then for every  $p \in X$ , the map  $\xi \mapsto H_\xi(p)$  is continuous in the following sense: if  $\xi_n \rightarrow \xi \in \partial X$  and  $K \subset X$  is compact, then  $H_{\xi_n}(p) \cap K \rightarrow H_\xi(p) \cap K$  uniformly in the Hausdorff topology.*

**Corollary 3.4.** *The functions  $(v, q) \mapsto b_v(q)$  and  $(\xi, p, q) \mapsto b_\xi(p, q)$  are continuous on  $SX \times X$  and  $\partial X \times X \times X$ , respectively.*

Now we can define the types of measures that we work with to carry out the counting argument.

**Definition 3.5.** *Given  $h > 0$ , an  $h$ -conformal density on  $\partial X$  is a family of finite measures  $\{\mu_p\}_{p \in X}$  on  $\partial X$  with the following properties.*

- (1)  $\text{supp } \mu_p = \partial X$  for all  $p \in X$ .
- (2)  $\{\mu_p\}_{p \in X}$  is  $\Gamma$ -equivariant: for all Borel sets  $A \subset \partial X$ , we have

$$\mu_{\gamma p}(\gamma A) = \mu_p(A).$$

- (3)  $\frac{d\mu_q}{d\mu_p}(\xi) = e^{-hb_\xi(q,p)}$  for almost all  $\xi \in \partial X$ .

In the setting of Theorem 1.2, an  $h$ -conformal density was obtained in [CKW19, Proposition 5.1] via a Patterson–Sullivan construction, where  $h$  is the topological entropy of the geodesic flow. Given a conformal density, the proof of [Kni98, Lemma 2.4] gives the following.

**Lemma 3.6.** *For  $p \in X$ , the measure  $\bar{\mu}$  on  $\partial^2 X \subset (\partial X)^2$  defined by*

$$(3.4) \quad d\bar{\mu}(\xi, \eta) = e^{h\beta_p(\xi, \eta)} d\mu_p(\xi) d\mu_p(\eta)$$

*is  $\Gamma$ -invariant.*

Under the hypotheses of Theorem 1.2, we proved in [CKW19, Theorem 5.6 and Remark 5.7] that  $\bar{\mu}$ -a.e. pair  $(\xi, \eta) \in \partial^2 X$  is connected by a unique geodesic  $V(\xi, \eta)$ , and that writing  $\lambda_{\xi, \eta}$  for Lebesgue measure along the trajectory corresponding to  $V(\xi, \eta)$ , the measure defined on  $SX$  by

$$(3.5) \quad m(A) = \int_{\partial^2 X} \lambda_{\xi, \eta}(A) d\bar{\mu}(\xi, \eta)$$

is related to the unique MME  $\underline{m}$  on  $SM$  by

$$(3.6) \quad m(A) = \int_{SM} \#(\text{pr}_*^{-1}(\underline{v}) \cap A) d\underline{m}(\underline{v}).$$

Moreover, in [CKW19, Theorem 6.1 and Remark 6.2] we used an argument of Babillot [Bab02] to prove that the flow is mixing with respect to  $\underline{m}$  under these same assumptions.

**3.2. Local product structure near expansive vectors.** From now on we fix an expansive vector  $v_0 \in \mathcal{E}$  (which is nonempty since it has full measure for  $m$ ), and let  $p = \pi(v_0) \in X$ . This will be a ‘reference point’ for all the definitions that follow, and we will suppress  $v_0$  and  $p$  from the notation. We fix a scale  $\epsilon \in (0, \min(\frac{1}{8}, \frac{\text{inj}(M)}{4})]$ , where  $\text{inj}(M)$  is the injectivity radius of  $M$ . At the very end of the proof, in §6.2, we will need to take a limit as  $\epsilon \rightarrow 0$ , but until then  $\epsilon$  will be fixed.

**Definition 3.7.** *The Hopf map  $H: SX \rightarrow \partial^2 X \times \mathbb{R}$  for  $p \in X$  is*

$$(3.7) \quad H(v) := (v^-, v^+, s(v)), \text{ where } s(v) := b_{v^-}(\pi v, p).$$

The following fundamental fact will be useful:

$$(3.8) \quad s(\phi^t v) = s(v) + t \text{ for all } v \in SX \text{ and } t \in \mathbb{R}.$$

Note that the Hopf map is continuous by Corollary 3.4 and the definition of the topology on  $\partial X$ . Following Ricks [Ric19], we use the Hopf map to define a local product structure on a neighborhood in  $SX$ . Given disjoint sets  $\mathbf{P}, \mathbf{F} \subset \partial X$ , the set  $H^{-1}(\mathbf{P} \times \mathbf{F} \times \{0\})$  represents the set of all  $v \in SX$  whose past history under  $\phi^t$  is given by  $\mathbf{P}$ , whose future evolution is given by  $\mathbf{F}$ , and such that  $b_{v^-}(\pi v, p) = 0$ .

**Lemma 3.8.** *For any disjoint closed sets  $\mathbf{P}, \mathbf{F} \subset \partial X$ , the set  $H^{-1}(\mathbf{P} \times \mathbf{F} \times \{0\}) \subset SX$  is compact.*

We will use the following choice of  $\mathbf{P}, \mathbf{F}$ : given our fixed choice of  $v_0 \in S_p X \cap \mathcal{E}$ , we consider for each  $\theta > 0$  the sets

$$(3.9) \quad \begin{aligned} \mathbf{P} &= \mathbf{P}_\theta := \{w^- : w \in S_p X \text{ and } \angle_p(w, v_0) \leq \theta\}, \\ \mathbf{F} &= \mathbf{F}_\theta := \{w^+ : w \in S_p X \text{ and } \angle_p(w, v_0) \leq \theta\}. \end{aligned}$$

The function  $\theta \mapsto \bar{\mu}(\mathbf{P}_\theta \times \mathbf{F}_\theta)$  is nondecreasing and thus has at most countably many discontinuities; from now on we assume that  $\theta$  is chosen to be a point of continuity of this function, so that

$$(3.10) \quad \lim_{\rho \rightarrow \theta} \bar{\mu}(\mathbf{P}_\rho \times \mathbf{F}_\rho) = \bar{\mu}(\mathbf{P}_\theta \times \mathbf{F}_\theta).$$

The assumption that  $v_0 \in \mathcal{E}$  lets us strengthen the boundedness result in Lemma 3.8.

**Lemma 3.9.** *Let  $v_0, p, \epsilon$  be as above. Then there exists  $\theta_1 > 0$  such that for all  $0 < \theta \leq \theta_1$  we have  $\text{diam } \pi H^{-1}(\mathbf{P} \times \mathbf{F} \times \{0\}) < \frac{\epsilon}{2}$ .*

Given  $\alpha \in (0, \frac{3}{2}\epsilon]$ , we consider the *flow box*

$$(3.11) \quad \begin{aligned} B &= B_\theta^\alpha := H^{-1}(\mathbf{P} \times \mathbf{F} \times [0, \alpha]) \\ &= \{w \in SX : (w^-, w^+) \in \mathbf{P} \times \mathbf{F} \text{ and } s(w) \in [0, \alpha]\} \\ &= \bigcup \{\phi^{[0, \alpha]} w : (w^-, w^+) \in \mathbf{P} \times \mathbf{F} \text{ and } s(w) = 0\}, \end{aligned}$$

which is a union of tangent vectors to geodesic rays of length  $\alpha$ ; note that each trajectory of  $\phi^t$  that enters  $B$  must do so through the “front face”  $H^{-1}(\mathbf{P} \times \mathbf{F} \times \{0\})$ .

We will occasionally write  $B^\alpha$  or  $B_\theta$  when only one of  $\alpha, \theta$  needs to be explicitly specified (the other being constant through a long portion of the proof). In fact we will only need to consider the cases  $\alpha = \epsilon$ ,  $\alpha = \epsilon \pm 4\epsilon^2$ , and  $\alpha = \epsilon^2$ . The last of these is important enough to deserve its own notation, and we write

$$(3.12) \quad S = S_\theta := B_\theta^{\epsilon^2} = H^{-1}(\mathbf{P} \times \mathbf{F} \times [0, \epsilon^2]).$$

Thus from now on we will always use  $B$  to denote a *box* with depth  $\alpha \approx \epsilon$ , and  $S$  to denote the *slice* with depth exactly  $\epsilon^2$ . We point out that since  $0 < \epsilon \leq \frac{1}{8}$ , we have  $\epsilon + 4\epsilon^2 \leq \frac{3}{2}\epsilon$  and  $\epsilon - 4\epsilon^2 \geq \frac{\epsilon}{2}$ . We also observe that by (3.10) and the product structure of  $S$  and  $\bar{B}$ , we have

$$(3.13) \quad \lim_{\rho \rightarrow \theta} m(S_\rho) = S_\theta \quad \text{and} \quad \lim_{\rho \rightarrow \theta} m(B_\rho^\alpha) = m(B_\theta^\alpha)$$

for every  $\alpha$ , and also

$$(3.14) \quad m(\partial B_\theta^\alpha) = 0.$$

The following is an immediate consequence of Lemma 3.9.

**Lemma 3.10.** *Let  $v_0, p, \epsilon$  be as above, and let  $\theta_1$  be as in Lemma 3.9. Then for all  $0 < \theta \leq \theta_1$  and  $\alpha \leq \frac{3}{2}\epsilon$ , we have  $\text{diam } \pi B_\theta^\alpha < 2\epsilon$ .*

The following lemma will be used in Lemmas 4.11 and 5.2; see Appendix A for the proof.

**Lemma 3.11.** *Given  $v_0, p, \epsilon > 0$  as above, there exists  $\theta_2 > 0$  such that for all  $0 < \theta \leq \theta_2$ , given any  $\xi, \eta \in \mathbf{P}_\theta$  and any  $q$  lying within  $2\epsilon$  of  $\pi H^{-1}(\mathbf{P}_\theta \times \mathbf{F}_\theta \times [0, \infty))$ , we have  $|b_\xi(q, p) - b_\eta(q, p)| < \epsilon^2$ , with a similar estimate when the roles of  $\mathbf{P}_\theta$  and  $\mathbf{F}_\theta$  are reversed.*

Let  $\theta_0 = \min(\theta_1, \theta_2)$ ; from now on we will always consider  $0 < \theta \leq \theta_0$  so that Lemmas 3.9, 3.10, and 3.11 all hold.

Using Lemma 3.10 and the fact that  $\epsilon < \frac{1}{4} \text{inj}(M)$ , we see that the quotient map  $\text{pr}: X \rightarrow M$  is injective on  $\pi B$ , and similarly  $\text{pr}_*: SX \rightarrow SM$  is injective on  $B$ . We will write

$$(3.15) \quad \underline{B} = \underline{B}_\theta^\alpha := \text{pr}_*(B_\theta^\alpha) \quad \text{and} \quad \underline{S} = \underline{S}_\theta := \text{pr}_*(S_\theta)$$

for the flow box and the slice in  $SM$ .

## 4. COUNTING WITH THE FUNDAMENTAL GROUP

**4.1. An upper counting bound.** In §2.4, we described the relationship between closed geodesics and deck transformations  $\gamma \in \Gamma$ . Now we return to this question, restricting our attention to closed geodesics that “pass through”  $\underline{B} \subset SM$  in the sense that  $\dot{c}(s) \in \underline{B}$  for some  $s$ . In particular, we estimate  $\#C(t)$  from Theorem 1.3 in terms of  $\nu_t(\underline{B})$  (recall (1.2)) and a certain subset of  $\Gamma$ . In §5 we will use the scaling

and mixing properties of  $\underline{m}$  to estimate the size of this subset; then in §6 we will show that  $\nu_t \rightarrow \underline{m}$  and combine all of these estimates to complete the proof.

To describe the subset of  $\Gamma$  that we work with, start by recalling from Theorem 1.3 that  $C(t)$  is any maximal set of pairwise non-free-homotopic closed geodesics in  $M$  with lengths in  $(t - \epsilon, t]$ , and that (1.2) gives

$$(4.1) \quad \#C(t) = \frac{\sum_{\underline{c} \in C(t)} \text{Leb}_{\underline{c}}(B_{\theta}^{\alpha})}{t\nu_t(B_{\theta}^{\alpha})} \text{ for every } \alpha, \theta.$$

By the definition of  $B_{\theta}^{\alpha}$ , every  $\underline{v} \in B_{\theta}^{\alpha}$  has the property that the connected component of 0 in  $\{s \in \mathbb{R} : \phi^s \underline{v} \in B_{\theta}^{\alpha}\}$  is an interval of length  $\alpha$ , which begins when  $\phi^s \underline{v} \in \text{pr}_* H^{-1}(\mathbf{P} \times \mathbf{F} \times \{0\})$ . Let  $\Pi(t)$  be the set of tangent vectors initiating one of these segments; that is,

$$(4.2) \quad \Pi(t) := \{\dot{\underline{c}}(s) \in \text{pr}_* H^{-1}(\mathbf{P} \times \mathbf{F} \times \{0\}) : \underline{c} \in C(t), s \in \mathbb{R}\}.$$

The set  $\Pi(t)$  is finite and does not depend on  $\alpha$ . From (4.1) we get

$$(4.3) \quad \#C(t) = \frac{\alpha \#\Pi(t)}{t \nu_t(B_{\theta}^{\alpha})}.$$

We estimate  $\#\Pi(t)$  by associating to each  $\underline{v} \in \Pi(t)$  the axial isometry of an appropriate lift.

**Definition 4.1.** *Given  $\underline{v} \in \Pi(t)$ , let  $\ell = \ell(\underline{v}) \in (t - \epsilon, t]$  be such that  $\phi^{\ell} \underline{v} = \underline{v}$ , and let  $v$  be the unique lift of  $\underline{v}$  such that  $v \in B_{\theta}^{\alpha}$ . Define  $\Theta(\underline{v}) \in \Gamma$  to be the axial isometry of  $(c_v, \ell)$ ; that is, the unique isometry of  $X$  such that  $\phi^{\ell} v = \Theta(v)_* v$ . Observe that  $|\Theta(v)| = \ell$ .*

Observe that for each  $\underline{v} \in \Pi(t)$  and  $\gamma = \Theta(\underline{v})$ , we have  $\phi^t v = \phi^{t-\ell} \gamma_* v \in \gamma_* B_{\theta}^{\epsilon}$ , and so  $v \in S_{\theta} \cap \phi^{-t} \gamma_* B_{\theta}^{\epsilon}$ . With this in mind, we define for each  $t > 0$  and  $\alpha \in (0, \frac{3}{2}\epsilon]$  the set

$$(4.4) \quad \Gamma(t, \alpha) = \Gamma_{\theta}(t, \alpha) := \{\gamma \in \Gamma : S_{\theta} \cap \phi^{-t} \gamma_* B_{\theta}^{\alpha} \neq \emptyset\}$$

and observe that  $\Theta(\Pi(t)) \subset \Gamma(t, \epsilon)$ . To relate  $\#\Pi(t)$  and  $\#\Gamma(t, \epsilon)$ , we need to control the multiplicity of  $\Theta$ .

In some instances, such as when  $M$  is an oriented surface, it can be shown that  $\Theta$  is injective. In general, injectivity may fail: for example, if  $\beta$  acts as translation once around a Möbius strip whose central circle has length  $t/2$ , then the boundary circle of length  $t$  will correspond to two elements of  $\Pi(t)$ , and  $\Theta$  maps each of these to  $\beta^2$ .

**Definition 4.2.** *Given  $\gamma \in \Gamma$ , let  $d = d(\gamma) \in \mathbb{N}$  be maximal such that  $\gamma = \beta^d$  for some  $\beta \in \Gamma$ .*

**Lemma 4.3.** *For every  $t$  and  $\gamma$ , we have  $\#\Theta^{-1}(\gamma) \leq d(\gamma)$ .*

*Proof.* Suppose  $\underline{v}, \underline{w} \in \Pi(t)$  have  $\Theta(\underline{v}) = \Theta(\underline{w}) =: \gamma$ . Then  $c_v$  and  $c_w$  are both axes of  $\gamma$ . By Lemma 2.12, the set of isometries fixing the

endpoints of  $c_v$  and  $c_w$  is an infinite cyclic group; let  $\beta$  be a generator of this group such that  $\gamma = \beta^{d(\gamma)}$ .

Lemma 2.17 implies that  $\underline{c}_v$  and  $\underline{c}_w$  are free-homotopic as curves of length  $|\gamma|$ . Since the elements of  $C(t)$  are pairwise non-free-homotopic, we conclude that  $\underline{c}_v$  and  $\underline{c}_w$  are reparametrizations of the same closed geodesic; in other words, there is  $s \in (0, |\gamma|]$  such that  $\phi^s(\underline{v}) = \underline{w}$ . Lifting gives  $\tau \in \Gamma$  such that  $\phi^s v = \tau_* w$ , and thus

$$\tau w^+ = \lim_{r \rightarrow \infty} \tau \pi \phi^r w = \lim_{r \rightarrow \infty} \pi \phi^{r+s} v = v^+ = w^+.$$

Similarly,  $\tau$  fixes  $w^-$ , so  $\tau = \beta^k$  for some  $k \geq 1$  (it must be positive since  $s > 0$ ), and moreover  $k \leq d(\gamma)$  because

$$\begin{aligned} k|\beta| &= |\tau| = d(\pi(\phi^s(v)), \pi w) \\ &\leq d(\pi(\phi^s(v)), \pi v) + d(\pi v, \pi w) \leq d(\gamma)|\beta| + \frac{\epsilon}{2} \leq (d(\gamma) + \frac{1}{2})|\beta|, \end{aligned}$$

where the last two inequalities use Lemma 3.9 and the fact that  $\epsilon \leq \frac{1}{4} \text{inj}(M) < |\beta|$  (by Lemma 2.15). Fixing  $\underline{v}$  we see that  $\underline{w}$  is uniquely determined by  $\tau$ , and since there are at most  $d(\gamma)$  choices for  $\tau$ , this proves the lemma.  $\square$

From Lemma 4.3, we deduce that

$$(4.5) \quad \#\Pi(t) \leq \#\Gamma(t, \epsilon) + \sum_{\substack{\gamma \in \Theta(\Pi(t)) \\ d(\gamma) \geq 2}} d(\gamma).$$

To use (4.3) and (4.5) to estimate  $\#C(t)$  in terms of  $\#\Gamma(t, \epsilon)$ , we need to estimate how many  $\gamma \in \Theta(\Pi(t))$  have larger values of  $d(\gamma)$ .

**Definition 4.4.** *Given disjoint closed sets  $\mathbf{Q}, \mathbf{R} \subset \partial X$ , consider for each  $d \geq 2$  and  $t > 0$  the set*

$$(4.6) \quad \Gamma_d(\mathbf{Q}, \mathbf{R}, t) := \{\gamma \in \Gamma : \xi_\gamma^- \in \mathbf{Q}, \xi_\gamma^+ \in \mathbf{R}, \\ d(\gamma) \geq d, \text{ and } |\gamma| \in (t - \epsilon, t]\},$$

where  $\xi_\gamma^\pm$  are as in Definition 2.7.

Observe that  $\{\gamma \in \Theta(\Pi(t)) : d(\gamma) \geq 2\} \subset \Gamma_2(\mathbf{P}, \mathbf{F}, t)$ .

**Lemma 4.5.** *Given any disjoint closed sets  $\mathbf{Q}, \mathbf{R} \subset \partial X$ , there is  $K > 0$  such that for all  $\epsilon \in (0, \text{inj } M]$  and  $t > 0$  we have*

$$\sum_{\gamma \in \Gamma_2(\mathbf{Q}, \mathbf{R}, t)} d(\gamma) \leq K e^{\frac{2}{3}ht}.$$

*Proof.* Let  $A := \pi H^{-1}(\mathbf{Q} \times \mathbf{R} \times \{0\})$ ; note that  $\text{diam } A < \infty$  by Lemma 3.8. Fix  $q \in A$ . Freire and Mañé proved in [FM82] that  $h = \lim_{r \rightarrow \infty} \frac{1}{r} \log \text{vol}(B(q, r))$ , where  $B(q, r)$  is the ball of radius  $r$  centered at  $q$  in  $X$ . Thus fixing  $\delta \in (0, h/3)$ , there is  $K_0 > 0$  such that

$$(4.7) \quad \text{vol}(B(q, r)) \leq K_0 e^{(h+\delta)r} \text{ for all } r > 0.$$

Given  $d \geq 2$  and  $\gamma \in \Gamma_d(\mathbf{Q}, \mathbf{R}, t)$ , write  $\gamma = \beta(\gamma)^{d(\gamma)}$ . Every axis of  $\beta(\gamma)$  is an axis of  $\gamma$ , and thus has endpoints  $\xi_\gamma^\pm$  by Lemma 2.10. In particular, we can choose  $v \in H^{-1}(\mathbf{Q} \times \mathbf{R} \times \{0\})$  tangent to such an axis, and observe that  $\phi^{|\beta(\gamma)|}v = \beta(\gamma)_*v$ , so that

$$\begin{aligned} d(q, \beta(\gamma)q) &\leq d(q, \pi v) + d(\pi v, \pi\beta(\gamma)_*v) + d(\pi\beta(\gamma)_*v, \beta(\gamma)q) \\ (4.8) \quad &\leq 2d(q, \pi v) + d(\pi v, \pi\phi^{|\beta(\gamma)|}v) \\ &\leq 2 \operatorname{diam} A + |\beta(\gamma)|. \end{aligned}$$

Observe that as  $\gamma$  ranges over  $\Gamma_d(\mathbf{Q}, \mathbf{R}, t)$ , the sets  $B(\beta(\gamma)q, \operatorname{inj} M)$  all have the same volume  $V$  (since deck transformations act isometrically) and are all disjoint because  $\gamma \mapsto \beta(\gamma)$  is injective (this uses our choice of  $\epsilon$  and the fact that  $|\beta| \geq 2 \operatorname{inj}(M) > \epsilon$  for all  $\beta \neq \operatorname{Id}$ ). Thus their union has volume  $V \cdot \#\Gamma_d(\mathbf{Q}, \mathbf{R}, t)$ . Since this union is contained in  $B(q, 2 \operatorname{diam} A + \frac{t}{d})$ , we can use (4.7) to deduce that

$$V \cdot \#\Gamma_d(\mathbf{Q}, \mathbf{R}, t) \leq \operatorname{vol}(B(q, 2 \operatorname{diam} A + \frac{t}{d})) \leq K_0 e^{(h+\delta)(\operatorname{diam} A + \frac{t}{d})}.$$

Writing  $K_1 = V^{-1}K_0 e^{(h+\delta) \operatorname{diam} A}$ , we have

$$(4.9) \quad \#\Gamma_d(\mathbf{Q}, \mathbf{R}, t) \leq K_1 e^{(h+\delta)t/d}.$$

Observe that  $d(\gamma) = |\gamma|/|\beta(\gamma)| \leq |\gamma|/\operatorname{inj}(M)$  by Lemma 2.15, so

$$\sum_{\gamma \in \Gamma_2(\mathbf{Q}, \mathbf{R}, t)} d(\gamma) = 2\#\Gamma_2(\mathbf{Q}, \mathbf{R}, t) + \sum_{3 \leq d \leq t/\operatorname{inj}(M)} \Gamma_d(\mathbf{Q}, \mathbf{R}, t).$$

From (4.9) we see that  $\#\Gamma_d(\mathbf{Q}, \mathbf{R}, t) \leq K_1 e^{(h+\delta)t/2}$  for each  $d \geq 2$ , and thus

$$\sum_{\gamma \in \Gamma_2(\mathbf{Q}, \mathbf{R}, t)} d(\gamma) \leq \frac{tK_1}{\operatorname{inj}(M)} e^{(h+\delta)t/2}.$$

Since  $(h + \delta)/2 < 2h/3$ , this proves the lemma.  $\square$

Combining Lemma 4.5 with (4.3) and (4.5) gives

$$(4.10) \quad \#C(t) \leq \frac{\epsilon}{tv_t(B^\epsilon)} (\#\Gamma(t, \epsilon) + Ke^{\frac{2}{3}ht}).$$

**Remark 4.6.** *Once we show that  $\lim_{t \rightarrow \infty} \frac{1}{t} \log \#\Gamma(t, \epsilon) = h$ , it will follow that the last term in (4.10) does not affect the asymptotics of  $\#C(t)$ . This same argument shows that the validity of (1.1) is not affected by whether we count all homotopy classes or only primitive ones; the number of nonprimitive ones is of a lower exponential order, and thus its contribution to the ratio of interest vanishes in the limit.*

**4.2. A type of closing lemma.** We will eventually complement the upper bound in (4.10) by getting a lower bound for  $\#C(t)$  in §4.4, but first we need a way to guarantee that the intersection  $S \cap \phi^{-t}\gamma_*B_\theta^\alpha$  actually contains a periodic orbit.

Recall from Lemma 2.15 that every  $\gamma \in \Gamma \setminus \operatorname{Id}$  has an axis  $c$ , which projects to a closed geodesic on  $M$ . To produce a set of  $\gamma$  for which

this axis passes through  $B$ , we need a condition guaranteeing that  $c(-\infty) \in \mathbf{P}$  and  $c(+\infty) \in \mathbf{F}$ , since then the tangents to  $c$  will pass through  $B$ . To this end we consider

$$(4.11) \quad \Gamma^* = \Gamma_\theta^* := \{\gamma \in \Gamma : \gamma\mathbf{F}_\theta \subset \mathbf{F}_\theta \text{ and } \gamma^{-1}\mathbf{P}_\theta \subset \mathbf{P}_\theta\}.$$

**Lemma 4.7.** *Given any  $\gamma \in \Gamma^*$ , there exists an axis  $c$  for  $\gamma$  such that  $c(-\infty) \in \mathbf{P}$  and  $c(+\infty) \in \mathbf{F}$ .*

*Proof.* By the Brouwer fixed point theorem,  $\gamma$  has one fixed point in  $\mathbf{P}$  and one in  $\mathbf{F}$ . By Lemma 2.15, it has an axis  $c$ . By Lemma 2.10, the fixed points of  $\gamma$  are the endpoints of  $c$ .  $\square$

This lemma guarantees that given any  $\gamma \in \Gamma^*$ , there is  $v \in S$  such that  $\phi^{|\gamma|}v = \gamma_*v$ . We will apply this in the situation when  $S \cap \gamma_*\phi^{-t}B \neq \emptyset$ ; in §4.3 we explore the relationship between  $t$  and  $|\gamma|$ . The result here, together with the control on the period established there, can be thought of as a type of closing lemma.

For the moment we establish a relationship between  $\Gamma_\rho(t)$  and  $\Gamma_\theta^*$  that will be important in §5.2, when we combine the lower and upper bounds into a single asymptotic estimate.

**Lemma 4.8.** *For every  $0 < \rho < \theta$ , there exists  $t_0 > 0$  such that for all  $t \geq t_0$  we have  $\Gamma_\rho(t) \subset \Gamma_\theta^*$ .*

To prove Lemma 4.8 we need the following consequence of the uniform visibility axiom.

**Lemma 4.9.** *Let  $\mathbf{Q}, \mathbf{R} \subset \partial X$  be disjoint compact sets. Fix a compact subset  $A \subset X$  and consider for each  $T > 0$  the set*

$$C_{\mathbf{Q}}^T := \{c_{p,\xi}(t) : p \in A, \xi \in \mathbf{Q}, t \geq T\},$$

where  $c_{p,\xi}$  is as in Lemma 2.5, and define  $C_{\mathbf{R}}^T$  similarly.

*Then given any open sets  $\mathbf{U}, \mathbf{V} \subset \partial X$  such that  $\mathbf{U} \supset \mathbf{Q}$  and  $\mathbf{V} \supset \mathbf{R}$ , there exists  $T > 0$  such that if  $\gamma \in \Gamma$  satisfies  $\gamma(C_{\mathbf{Q}}^T) \cap A \neq \emptyset$  and  $\gamma(A) \cap C_{\mathbf{R}}^T \neq \emptyset$ , then  $\gamma(\partial X \setminus \mathbf{U}) \subset \mathbf{V}$  and  $\gamma^{-1}(\partial X \setminus \mathbf{V}) \subset \mathbf{U}$ .*

*Proof.* We prove the first inclusion; the second is similar. Choose  $\epsilon > 0$  sufficiently small that for every  $p \in A$ ,  $\xi \in \mathbf{R}$ , and  $\eta \in \partial X$  with  $\angle_p(\xi, \eta) < 2\epsilon$ , we have  $\eta \in \mathbf{V}$ , and similarly with  $\mathbf{R}, \mathbf{V}$  replaced by  $\mathbf{Q}, \mathbf{U}$ . (Existence of such an  $\epsilon$  for each individual  $p$  is immediate from the definition of the topology on  $\partial X$ ; the fact that  $\epsilon$  can be chosen independently of  $p$  uses compactness of  $A$  and continuity of the map  $p \mapsto \dot{c}_{p,\xi}(0)$  from Lemma 2.5.) Let  $L = L(\epsilon)$  be given by the uniform visibility property (see Definition 2.1), and let  $T = 2L + \text{diam } A$ .

Fix  $\gamma$  satisfying the hypothesis of the lemma, and let  $x, y \in A$  be such that  $\gamma y \in C_{\mathbf{R}}^T$  and  $\gamma^{-1}x \in C_{\mathbf{Q}}^T$ . By definition of  $C_{\mathbf{R}}^T$ , there is  $p \in A$  such that  $\xi := c_{p,\gamma y}(\infty) \in \mathbf{R}$  and  $d(p, \gamma y) \geq T$ . Similarly, there is  $q \in A$  such that  $c_{q,\gamma^{-1}x}(\infty) \in \mathbf{Q}$  and  $d(q, \gamma^{-1}x) \geq T$ .

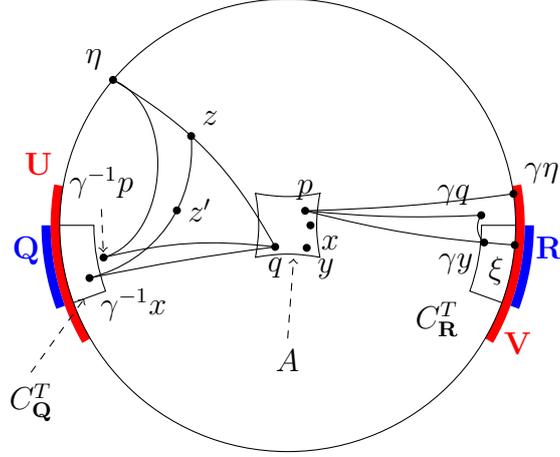


FIGURE 2. Proving Lemma 4.9

Given an arbitrary  $\eta \in \partial X \setminus \mathbf{U}$ , we will show that  $\angle_p(\xi, \gamma\eta) < 2\epsilon$ . Observe that

$$(4.12) \quad \angle_p(\xi, \gamma\eta) \leq \angle_p(\xi, \gamma q) + \angle_p(\gamma\eta, \gamma q) = \angle_p(\gamma y, \gamma q) + \angle_{\gamma^{-1}p}(\eta, q).$$

We bound  $\angle_p(\gamma y, \gamma q)$  by observing that  $d(\gamma y, \gamma q) = d(y, q) \leq \text{diam } A$ , while  $d(p, \gamma y) \geq T$ . Thus for every point  $y'$  on the geodesic segment connecting  $\gamma y$  and  $\gamma q$ , we have  $d(p, y') \geq T - \text{diam } A \geq L$ , and the uniform visibility property implies that

$$(4.13) \quad \angle_p(\gamma y, \gamma q) < \epsilon.$$

To bound  $\angle_{\gamma^{-1}p}(\eta, q)$ , first note by our choice of  $\epsilon$  and  $q$ , for every  $z$  on the geodesic ray from  $q$  to  $\eta$  we have  $\angle_q(\gamma^{-1}x, z) = \angle_q(\gamma^{-1}x, \eta) \geq 2\epsilon$ . By the uniform visibility property, for each such  $z$  there is  $z'$  on the geodesic segment from  $z$  to  $\gamma^{-1}x$  such that  $d(q, z') \leq L$ . Note also that

$$d(z, \gamma^{-1}x) \geq d(z', \gamma^{-1}x) \geq d(q, \gamma^{-1}x) - d(q, z') \geq T - L,$$

and thus

$$d(z, \gamma^{-1}p) \geq d(z, \gamma^{-1}x) - d(\gamma^{-1}x, \gamma^{-1}p) \geq T - L - \text{diam } A \geq L$$

since  $\gamma^{-1}$  acts isometrically. Applying the uniform visibility property once more gives  $\angle_{\gamma^{-1}p}(\eta, q) < \epsilon$ . Together with (4.12) and (4.13) this gives  $\angle_p(\xi, \gamma\eta) < 2\epsilon$ , and thus  $\gamma\eta \in \mathbf{V}$ , which proves the lemma.  $\square$

*Proof of Lemma 4.8.* We apply Lemma 4.9 by putting  $\mathbf{Q} = \mathbf{P}_\rho$ ,  $\mathbf{R} = \mathbf{F}_\rho$ ,  $A = \pi B_\rho$ , and letting  $\mathbf{U}, \mathbf{V}$  be the interiors of  $\mathbf{P}_\theta, \mathbf{F}_\theta$ , respectively. Let  $T$  be given by that lemma; then for every  $t \geq T$  and  $\gamma \in \Gamma_\rho(t)$ , there exists  $v \in S_\rho \subset B_\rho$  such that  $w := \gamma_*^{-1}\phi^t v \in B_\rho$ . Putting  $x = \pi v \in A$  we have

$$\gamma^{-1}x = \pi \gamma_*^{-1}v = \pi \phi^{-t}w;$$

since  $\pi w \in A$ ,  $w^- \in \mathbf{P}_\rho = \mathbf{Q}$ , and  $t \geq T$ , this implies that  $\gamma^{-1}x \in C_{\mathbf{Q}}^T$ . Similarly, putting  $y = \pi w \in A$  gives

$$\gamma y = \pi \gamma_* w = \pi \phi^t v;$$

since  $\pi v \in A$ ,  $v^+ \in \mathbf{F}_\rho = \mathbf{R}$ , and  $t \geq T$ , this gives  $\gamma y \in C_{\mathbf{R}}^T$ , and Lemma 4.9 yields the desired result.  $\square$

**4.3. Depths of intersections.** Recall that we fix  $v_0 \in \mathcal{E}$  and  $p = \pi v_0$ . Consider the set

$$(4.14) \quad \Gamma_\theta^*(t, \alpha) := \Gamma^* \cap \Gamma_\theta(t, \alpha) = \{\gamma \in \Gamma^* : S_\theta \cap \gamma_* \phi^{-t} B_\theta^\alpha \neq \emptyset\}.$$

We need to understand how  $|\gamma|$  and  $t$  are related when  $\gamma \in \Gamma^*(t, \alpha)$ .

**Definition 4.10.** Given  $\xi \in \partial X$  and  $\gamma \in \Gamma$ , let  $b_\xi^\gamma := b_\xi(\gamma p, p)$ .

**Lemma 4.11.** For all  $\xi, \eta \in \mathbf{P}$  and every  $\gamma \in \Gamma_\theta(t, \alpha)$  with  $t > 0$ , we have  $|b_\xi^\gamma - b_\eta^\gamma| < \epsilon^2$ .

*Proof.* Given any such  $\gamma$ , there exists  $v \in S \cap \gamma_* \phi^{-t} B^\alpha$ , so  $\phi^t v \in \gamma_* B^\alpha$ . Thus there is  $q \in \pi B^\alpha$  such that  $\gamma q = \pi \phi^t v \in \pi H^{-1}(\mathbf{P}_\theta \times \mathbf{F}_\theta \times [0, \infty))$ . Since  $p \in \pi B^\alpha$ , we have  $d(\gamma p, \gamma q) = d(p, q) < 2\epsilon$  by Lemma 3.10, so  $\gamma p$  satisfies the condition of Lemma 3.11, and we have  $|b_\xi(\gamma p, p) - b_\eta(\gamma p, p)| < \epsilon^2$  for all  $\xi, \eta \in \mathbf{P}_\theta$ .  $\square$

**Lemma 4.12.** If  $c$  is an axis of  $\gamma$  and  $\xi = c(-\infty)$ , then  $b_\xi^\gamma = |\gamma|$ .

*Proof.* Let  $q$  be any point on  $c$ ; then  $|\gamma| = d(\gamma q, q) = b_\xi(\gamma q, q)$ . Thus (3.1) gives

$$\begin{aligned} |\gamma| - b_\xi^\gamma &= b_\xi(\gamma q, q) - b_\xi(\gamma p, p) \\ &= (b_\xi(\gamma q, \gamma p) + b_\xi(\gamma p, q)) - (b_\xi(\gamma p, q) + b_\xi(q, p)) \\ &= b_\xi(\gamma q, \gamma p) - b_{\gamma\xi}(\gamma q, \gamma p), \end{aligned}$$

where the last step uses the fact that  $\gamma$  acts isometrically so Busemann functions are unchanged when  $\gamma$  is applied to all three arguments. Finally,  $\gamma\xi = \xi$ , so the lemma is proved.  $\square$

**Lemma 4.13.** Given any  $\gamma \in \Gamma^*$  and any  $t \in \mathbb{R}$ , we have<sup>13</sup>

$$S \cap \phi^{-t} \gamma_* B^\alpha = \{w \in E^{-1}(\mathbf{P} \times \gamma \mathbf{F}) : s(w) \in [0, \epsilon^2] \cap (b_{w^-}^\gamma - t + [0, \alpha])\}.$$

*Proof.* To prove that  $S \cap \phi^{-t} \gamma_* B^\alpha \subset E^{-1}(\mathbf{P} \times \gamma \mathbf{F})$ , we observe that if  $E(w) \notin \mathbf{P} \times \gamma \mathbf{F}$ , then either  $w^- \notin \mathbf{P}$ , so  $w \notin S$ , or  $w^+ \notin \gamma \mathbf{F}$ , so  $w \notin \phi^{-t} \gamma_* B^\alpha$ .

<sup>13</sup>The asymmetry in this lemma – the fact that  $w^-$  appears and  $w^+$  does not – is due to the fact that we define  $s(w)$  using  $w^-$ ; in the original proof by Margulis, one must similarly choose whether to construct the front of the flow box as a union of stable leaves or unstable leaves. Of course either choice leads to a similar argument.

It remains to show that given  $w \in E^{-1}(\mathbf{P} \times \gamma\mathbf{F})$ , we have

$$(4.15) \quad w \in S \Leftrightarrow s(w) \in [0, \epsilon^2], \text{ and}$$

$$(4.16) \quad w \in \phi^{-t}\gamma_*B^\alpha \Leftrightarrow s(w) \in b_{w^-}^\gamma - t + [0, \alpha].$$

The first of these is immediate from the definition of  $S$ . For the second, we observe that  $s(v) = b_{v^-}(\pi v, p) = b_{\gamma v^-}(\gamma\pi v, \gamma p)$ , and thus

$$\begin{aligned} \gamma_*B^\alpha &= \{\gamma_*v : v \in E^{-1}(\mathbf{P} \times \mathbf{F}) \text{ and } b_{v^-}(\pi v, p) \in [0, \alpha]\} \\ &= \{w \in E^{-1}(\gamma\mathbf{P} \times \gamma\mathbf{F}) : b_{w^-}(\pi w, \gamma p) \in [0, \alpha]\} \end{aligned}$$

By (3.1) and (3.2), we have

$$b_{w^-}(\pi w, \gamma p) = b_{w^-}(\pi w, p) + b_{w^-}(p, \gamma p) = s(w) - b_{w^-}^\gamma;$$

moreover, since  $s(\phi^t w) = s(w) + t$  by (3.8), we see that  $\phi^t w \in \gamma_*B^\alpha$  if and only if  $s(w) - b_{w^-}^\gamma + t \in [0, \alpha]$ , which proves (4.16) and completes the proof of the lemma.  $\square$

**Lemma 4.14.** *If  $\gamma \in \Gamma^*(t, \alpha)$ , then  $|\gamma| \in [t - \alpha - \epsilon^2, t + 2\epsilon^2]$ .*

*Proof.* Since  $S \cap \phi^{-t}\gamma_*B^\alpha$  is nonempty, Lemma 4.13 guarantees that there exist  $\zeta \in \mathbf{P}$  and  $\tau \in [0, \epsilon^2]$  such that  $\tau \in b_\zeta^\gamma - t + [0, \alpha]$ ; this implies that

$$b_\zeta^\gamma \in t + \tau - [0, \alpha] \subset [t - \alpha, t + \epsilon^2].$$

From Lemmas 4.11 and 4.12, we have  $|\gamma| \in [b_\zeta^\gamma - \epsilon^2, b_\zeta^\gamma + \epsilon^2]$ , which proves the lemma.  $\square$

**4.4. A lower counting bound.** Now we can obtain a lower bound for  $\#\Pi(t) = \#\Theta(\Pi(t))$ , and hence for  $\#C(t)$ . We define

$$(4.17) \quad \Gamma'(t, \alpha) := \{\gamma \in \Gamma^*(t, \alpha) : \gamma \neq \beta^n \text{ for any } \beta \in \Gamma, n \geq 2\}.$$

**Lemma 4.15.** *Let  $\alpha = \epsilon - 4\epsilon^2$ ; then  $\Theta(\Pi(t)) \supset \Gamma'(t - 2\epsilon^2, \alpha)$ , and from (4.3) we obtain*

$$(4.18) \quad \#C(t) \geq \frac{\alpha}{t} \cdot \frac{\#\Gamma'(t - 2\epsilon^2, \alpha)}{\nu_t(\underline{B}^\alpha)}.$$

*Proof.* By Lemma 4.7, given any  $\gamma \in \Gamma'(t - 2\epsilon^2, \alpha)$ , there exists  $v \in H^{-1}(\mathbf{P} \times \mathbf{F} \times \{0\})$  such that  $\phi^{|\gamma|}v = \gamma_*v$ . By Lemma 4.14, we have

$$\begin{aligned} |\gamma| &\geq (t - 2\epsilon^2) - \alpha - \epsilon^2 = t - 3\epsilon^2 - (\epsilon - 4\epsilon^2) > t - \epsilon, \\ |\gamma| &\leq (t - 2\epsilon^2) + 2\epsilon^2 = t, \end{aligned}$$

and thus  $\underline{v} := \text{pr}_*v \in SM$  lies on a closed geodesic  $\underline{c}_v$  with length  $|\gamma| \in (t - \epsilon, t]$  (this uses the assumption that  $\gamma$  is not a nontrivial power of another isometry). By Lemmas 2.16 and 2.17, every closed geodesic in the free homotopy class of  $\underline{c}_v$  has the same period, so we choose  $\underline{c} \in C(t)$  that is homotopic to  $\underline{c}_v$ , and note from Lemma 2.16 that there is a lift  $c$  with  $w := \dot{c}(0) \in H^{-1}(\mathbf{P} \times \mathbf{F} \times \{0\})$ . It follows from (4.2) that  $\underline{w} := \text{pr}_*w \in \Pi(t)$ , and from Definition 4.1 and irreducibility of  $\gamma$  that  $\Theta(w) = \gamma$ , which proves the lemma.  $\square$

Observe that  $\Gamma^*(t - 2\epsilon^2, \alpha) \setminus \Gamma'(t - 2\epsilon^2, \alpha) \subset \Gamma_2(\mathbf{P}, \mathbf{F}, t)$ , so by Lemma 4.5 we have

$$\#\Gamma^*(t - 2\epsilon^2, \alpha) - \#\Gamma'(t - 2\epsilon^2, \alpha) \leq Ke^{\frac{2}{3}ht}.$$

Using this together with Lemma 4.15 gives

$$(4.19) \quad \#C(t) \geq \frac{\alpha}{t\nu_t(\underline{B}^\alpha)} (\#\Gamma^*(t - 2\epsilon^2, \alpha) - Ke^{\frac{2}{3}ht}), \quad \alpha = \epsilon - 4\epsilon^2.$$

From (4.10) and (4.19) we see that we must now estimate  $\#\Gamma(t, \alpha)$  and  $\#\Gamma^*(t, \alpha)$ ; we do this in the next section. Then in §6 we use these estimates to deduce the equidistribution result (Theorem 1.3) and combine them with the lemmas from this section to obtain good estimates on  $\#C(t)$ . This yields estimates on  $\#P(t)$  via Riemann sums, and sending  $\epsilon \rightarrow 0$  yields integrals that we can evaluate to prove Theorem 1.2.

## 5. CONSEQUENCES OF SCALING AND MIXING

In what follows, it will be convenient to use the following notations, along with  $f \sim g$ :

$$f(t) = e^{\pm C}g(t) \quad \Leftrightarrow \quad e^{-C}g(t) \leq f(t) \leq e^Cg(t) \text{ for all } t;$$

$$f(t) \lesssim g(t) \quad \Leftrightarrow \quad \limsup_{t \rightarrow \infty} \frac{f(t)}{g(t)} \leq 1;$$

$$f(t) \gtrsim g(t) \quad \Leftrightarrow \quad \liminf_{t \rightarrow \infty} \frac{f(t)}{g(t)} \geq 1;$$

$$f(t) \sim e^{\pm C}g(t) \quad \Leftrightarrow \quad e^{-C}g(t) \lesssim f(t) \lesssim e^Cg(t).$$

**5.1. Scaling.** Given  $\alpha \leq \frac{3}{2}\epsilon$ , we will estimate  $\#\Gamma(t, \alpha)$  and  $\#\Gamma^*(t, \alpha)$  using the product structure, scaling properties, and mixing property of  $m$ . Note from Lemma 4.13 that although  $B$  and  $S$  have a product structure given by (3.11) and (3.12), the sets  $S \cap \phi^{-t}\gamma_*B^\alpha$  do not always have such a structure. Using the formula in that lemma, though, we can give a sufficient condition for these intersections to have a product structure.

**Lemma 5.1.** *Given any  $\alpha, t > 0$  and  $\gamma \in \Gamma^*(t, \alpha)$ , we have*

$$S \cap \phi^{-(t+2\epsilon^2)}\gamma_*B^{\alpha+4\epsilon^2} = H^{-1}(\mathbf{P} \times \gamma\mathbf{F} \times [0, \epsilon^2]) =: S^\gamma.$$

*Proof.* By Lemma 4.13, the fact that  $S \cap \phi^{-t}\gamma_*B^\alpha \neq \emptyset$  implies existence of  $\eta \in \mathbf{P}$  such that

$$(b_\eta^\gamma - t + [0, \alpha]) \cap [0, \epsilon^2] \neq \emptyset,$$

from which we deduce that

$$b_\eta^\gamma - t - \epsilon^2 + [0, \alpha + 2\epsilon^2] \supset [0, \epsilon^2].$$

By Lemma 4.11, it follows that every  $\xi \in \mathbf{P}$  has

$$(b_\xi^\gamma - t - \epsilon^2 + [0, \alpha + 2\epsilon^2]) \cap [0, \epsilon^2] \neq \emptyset,$$

which in turn implies that

$$b_\xi^\gamma - t - 2\epsilon^2 + [0, \alpha + 4\epsilon^2] \supset [0, \epsilon^2].$$

By Lemma 4.13, this completes the proof.  $\square$

Given  $\gamma \in \Gamma^*$ , let  $S^\gamma := H^{-1}(\mathbf{P} \times \gamma\mathbf{F} \times [0, \epsilon^2])$  as in Lemma 5.1, and write  $\underline{S}^\gamma = \text{pr}_* S^\gamma \subset SM$ .

**Lemma 5.2.** *For each  $\gamma \in \Gamma^*$ , we have*

$$m(S^\gamma) = e^{\pm 2h\epsilon} e^{-h|\gamma|} m(S),$$

and similarly with  $m, S, S^\gamma$  replaced by  $\underline{m}, \underline{S}, \underline{S}^\gamma$ .

*Proof.* Since  $\underline{m}(\underline{S}^\gamma) = m(S^\gamma) = \epsilon^2 \bar{\mu}(\mathbf{P} \times \gamma\mathbf{F})$ , and  $\underline{m}(\underline{S}) = m(S) = \epsilon^2 \bar{\mu}(\mathbf{P} \times \mathbf{F})$ , it suffices to show that  $\bar{\mu}(\mathbf{P} \times \gamma\mathbf{F}) = e^{\pm 2h\epsilon} e^{-h|\gamma|} \bar{\mu}(\mathbf{P} \times \mathbf{F})$ .

We will use the definition of  $\bar{\mu}$  in (3.4) and the scaling properties of the conformal measure in Definition 3.5; thus we need to control  $\beta_p(\xi, \eta)$  for  $(\xi, \eta) \in \mathbf{P} \times \mathbf{F}$ , and  $b_\eta(\gamma^{-1}p, p)$  for  $\eta \in \mathbf{F}$ . The former will be close to 0, and the latter will be close to  $|\gamma|$ .

Indeed, given  $(\xi, \eta) \in \mathbf{P} \times \mathbf{F}$ , we can take  $q$  to lie on a geodesic connecting  $\xi$  and  $\eta$ , with  $b_\xi(q, p) = 0$ ; then (3.3) gives

$$|\beta_p(\xi, \eta)| = |b_\xi(q, p) + b_\eta(q, p)| \leq d(q, p) < \epsilon/2,$$

where the last inequality uses Lemma 3.9. Using this together with (3.4) gives

$$\bar{\mu}(\mathbf{P} \times \mathbf{F}) = e^{\pm h\epsilon/2} \mu_p(\mathbf{P}) \mu_p(\mathbf{F}) \text{ and } \bar{\mu}(\mathbf{P} \times \gamma\mathbf{F}) = e^{\pm h\epsilon/2} \mu_p(\mathbf{P}) \mu_p(\gamma\mathbf{F}),$$

and thus

$$(5.1) \quad \frac{\bar{\mu}(\mathbf{P} \times \gamma\mathbf{F})}{\bar{\mu}(\mathbf{P} \times \mathbf{F})} = e^{\pm h\epsilon} \frac{\mu_p(\gamma\mathbf{F})}{\mu_p(\mathbf{F})}.$$

We will estimate the latter ratio using Definition 3.5, whose Property 2 gives

$$\mu_p(\gamma\mathbf{F}) = \mu_{\gamma^{-1}p}(\mathbf{F}),$$

and whose Property 3 gives

$$\frac{d\mu_{\gamma^{-1}p}}{d\mu_p}(\eta) = e^{-hb_\eta(\gamma^{-1}p, p)}.$$

When  $\eta = c(-\infty)$ , where  $c$  is the axis for  $\gamma^{-1}$ , we have  $\eta \in \mathbf{F}$  because  $\gamma \in \Gamma^*$ , and  $b_\eta(\gamma^{-1}p, p) = |\gamma^{-1}| = |\gamma|$  by Lemma 4.12. For other choices of  $\eta \in \mathbf{F}$ , Lemma 3.11 implies that the value of  $b_\eta(\gamma^{-1}p, p)$  varies by at most  $\epsilon^2$ . We conclude that  $\mu_p(\gamma\mathbf{F}) = e^{\pm \epsilon^2} e^{-h|\gamma|} \mu_p(\mathbf{F})$ , and together with (5.1) this proves the lemma.  $\square$

From Lemmas 4.14 and 5.2 we immediately deduce the following.

**Corollary 5.3.** *Given  $\alpha \leq \frac{3}{2}\epsilon$  and  $\gamma \in \Gamma^*(t, \alpha)$ , we have  $|t - |\gamma|| \leq 2\epsilon$ , and thus  $m(S^\gamma) = e^{\pm 4h\epsilon} m(S)$ , and similarly for  $\underline{m}, \underline{S}, \underline{S}^\gamma$ .*

**5.2. Asymptotic estimates.** It follows from Lemmas 4.8 and 5.1 that given any  $\alpha \in (0, \frac{3}{2}\epsilon]$  and  $\rho \in (0, \theta)$ , for all sufficiently large  $t$  we have

$$\underline{S}_\rho \cap \phi^{-t} \underline{B}_\rho^\alpha \subset \bigcup_{\gamma \in \Gamma^*(t, \alpha)} \underline{S}_\theta^\gamma \subset \underline{S}_\theta \cap \phi^{-(t+2\epsilon^2)} \underline{B}_\theta^{\alpha+4\epsilon^2}.$$

Using Corollary 5.3 gives

$$(5.2) \quad \underline{m}(\underline{S}_\theta^\gamma) = e^{\pm 4h\epsilon} e^{-ht} \underline{m}(\underline{S}_\theta)$$

for all  $\gamma \in \Gamma^*(t)$ , and thus

$$\begin{aligned} e^{-4h\epsilon} \underline{m}(\underline{S}_\rho \cap \phi^{-t} \underline{B}_\rho^\alpha) &\leq \#\Gamma^*(t, \alpha) e^{-ht} \underline{m}(\underline{S}_\theta) \\ &\leq e^{4h\epsilon} \underline{m}(\underline{S}_\theta \cap \phi^{-(t+2\epsilon^2)} \underline{B}_\theta^{\alpha+4\epsilon^2}). \end{aligned}$$

Sending  $t \rightarrow \infty$ , using mixing, and dividing through by  $\underline{m}(\underline{S}_\theta) \underline{m}(\underline{B}_\theta^\alpha) = m(S_\theta) m(B_\theta^\alpha)$ , we get

$$e^{-4h\epsilon} \frac{m(S_\rho) m(B_\rho^\alpha)}{m(S_\theta) m(B_\theta^\alpha)} \lesssim \frac{\#\Gamma^*(t, \alpha)}{e^{ht} m(B_\theta^\alpha)} \lesssim e^{4h\epsilon} \frac{m(B_\theta^{\alpha+4\epsilon^2})}{m(B_\theta^\alpha)}$$

By (3.13),  $\theta$  is a point of continuity for  $\rho \mapsto m(S_\rho)$  and  $\rho \mapsto m(B_\rho^\alpha)$ , so we can send  $\rho \nearrow \theta$  and obtain

$$(5.3) \quad e^{-4h\epsilon} \lesssim \frac{\#\Gamma^*(t, \alpha)}{e^{ht} m(B_\theta^\alpha)} \lesssim e^{4h\epsilon} (1 + 4\epsilon^2/\alpha).$$

We will also need to use (5.3) with  $\Gamma^*$  replaced by  $\Gamma$ . Observe that for every  $\rho > \theta$ , Lemma 4.8 gives  $\Gamma_\theta^*(t, \alpha) \subset \Gamma_\theta(t, \alpha) \subset \Gamma_\rho^*(t, \alpha)$  for all sufficiently large  $t$ , and thus (5.3) gives

$$\begin{aligned} e^{-4h\epsilon} e^{ht} m(B_\theta^\alpha) &\lesssim \#\Gamma_\theta^*(t, \alpha) \lesssim \#\Gamma_\theta(t, \alpha) \\ &\lesssim \#\Gamma_\rho^*(t, \alpha) \lesssim e^{4h\epsilon} (1 + 4\epsilon^2/\alpha) m(B_\rho^\alpha) e^{ht} \end{aligned}$$

Sending  $\rho \searrow \theta$  and using (3.13) gives

$$(5.4) \quad e^{-4h\epsilon} e^{ht} m(B_\theta^\alpha) \lesssim \#\Gamma_\theta(t, \alpha) \lesssim e^{4h\epsilon} (1 + 4\epsilon^2/\alpha) e^{ht} m(B_\theta^\alpha).$$

## 6. COMPLETION OF THE PROOFS

**6.1. Equidistribution to the measure of maximal entropy.** The following is a standard result in ergodic theory; see for example the proof of [Wal82, Theorem 9.10].

**Lemma 6.1.** *Let  $Y$  be a compact metric space and  $\phi^t$  a continuous flow on  $Y$ . Fix  $\epsilon > 0$  and suppose that  $E_t \subset Y$  is a  $(t, \epsilon)$ -separated set for all sufficiently large  $t$ . Then the measures  $\mu_t$  defined by*

$$(6.1) \quad \mu_t(A) := \frac{1}{\#E_t} \sum_{v \in E_t} \frac{1}{t} \int_0^t \chi_A(\phi^s v) ds$$

have the property that if  $t_k \rightarrow \infty$  and the weak\* limit  $\mu = \lim_{k \rightarrow \infty} \mu_{t_k}$  exists, then  $h_\mu(\phi^1) \geq \limsup_{k \rightarrow \infty} \frac{1}{t_k} \log \#E_{t_k}$ .

In particular, if  $\lim_{t \rightarrow \infty} \frac{1}{t} \log \#E_t$  is the topological entropy of the flow, then every weak\* limit point of the family of measures  $\mu_t$  is a measure of maximal entropy.

**Corollary 6.2.** *Let  $Y$  be a compact metric space and  $\phi^t$  a continuous flow on  $Y$  with a unique measure of maximal entropy  $\mu$ . Fix  $\epsilon > 0$ . If  $\{E_t \subset Y\}_t$  is a family of  $(t, \epsilon)$ -separated sets for which  $\lim_{t \rightarrow \infty} \frac{1}{t} \log \#E_t$  is the topological entropy of the flow, then the measures  $\mu_t$  defined in (6.1) converge to  $\mu$  in the weak\* topology.*

*Proof of Theorem 1.3.* Let  $C(t)$  be as in the statement of Theorem 1.3. By (4.19), we have

$$\#C(t) \geq \frac{\alpha}{t} (\#\Gamma^*(t - 2\epsilon^2, \alpha) - Ke^{\frac{2}{3}ht}), \quad \alpha = \epsilon - 4\epsilon^2,$$

and then (5.3) gives

$$(6.2) \quad \varliminf_{t \rightarrow \infty} \frac{1}{t} \log \#C(t) \geq h.$$

We claim that the set  $\{\underline{c}(0) : \underline{c} \in C(t)\}$  is  $(t, \epsilon)$ -separated for any  $\epsilon \in (0, \text{inj}(M))$ . Indeed, if it were not, then  $C(t)$  would contain  $\underline{c}_0 \neq \underline{c}_1$  such that  $d(\underline{c}_0(t), \underline{c}_1(t)) \leq \text{inj}(M)$  for all  $t \in \mathbb{R}$ , and thus there would be lifts  $c_0, c_1$  satisfying the same inequality, so that  $c_0(\pm\infty) = c_1(\pm\infty)$ . By Lemma 2.12 there is  $\gamma \in \Gamma$  such that the axial isometries of  $c_0, c_1$  are both of the form  $\gamma^k$ . Since  $\epsilon < \text{inj}(M) < |\gamma|$ , the value of  $k$  is the same for both geodesics, so Lemma 2.17 implies that  $\underline{c}_0$  and  $\underline{c}_1$  are homotopic, contradicting our assumption.

We conclude that the limit in (6.2) exists and is equal to  $h$ , and Corollary 6.2 implies that the measures  $\nu_t$  from (1.2) converge to the unique measure of maximal entropy, which completes the proof of Theorem 1.3.  $\square$

By Theorem 1.3 and (3.14), we have  $\nu_t(\underline{B}^\epsilon) \rightarrow \underline{m}(\underline{B}^\epsilon) = m(B^\epsilon)$ . Now (4.10) and (5.4) give

$$(6.3) \quad \#C(t) \lesssim \frac{\epsilon}{t} \cdot \frac{\#\Gamma(t, \epsilon) + Ke^{\frac{2}{3}ht}}{m(B^\epsilon)} \lesssim e^{4h\epsilon} (1 + 4\epsilon) \frac{\epsilon}{t} e^{ht}.$$

Similarly, (4.19) and (5.3) (with  $\alpha = \epsilon - 4\epsilon^2$ ) give

$$\#C(t) \gtrsim \frac{\epsilon - 4\epsilon^2}{t} \cdot \frac{\#\Gamma^*(t - 2\epsilon^2, \alpha) - Ke^{\frac{2}{3}ht}}{m(B^\alpha)} \gtrsim (1 - 4\epsilon) e^{-4h\epsilon} \frac{\epsilon}{t} e^{ht} e^{-2h\epsilon^2}.$$

Combining this with (6.3) gives

$$(6.4) \quad \#C(t) \sim e^{\pm Q\epsilon} \frac{\epsilon}{t} e^{ht}$$

where  $Q$  is a universal constant depending only on  $h$ .

**6.2. Riemann sums and integrals.** To complete the proof, we use (6.4) to estimate  $\#P(t)$  via a Riemann sum, and then send  $\epsilon \rightarrow 0$  to convert this to an integral and send the size of the error to 0. See [Ric19, §13] for a more axiomatic approach to this step.

Since  $\#P(T) \rightarrow \infty$  as  $T \rightarrow \infty$ , and each  $\#P(b)$  is finite, we see that  $\#P(T) \sim \#(P(T) \setminus P(b))$  for every  $b > 0$ . Summing over  $t_k = T - k\epsilon$  gives

$$(6.5) \quad \sum_{k=0}^{\lfloor (T-b)/\epsilon \rfloor} \#C(t_k) \leq \#(P(T) \setminus P(b)) \leq \sum_{k=0}^{\lceil (T-b)/\epsilon \rceil} \#C(t_k).$$

By (6.4) we can choose  $b$  sufficiently large that  $\#C(t) = e^{\pm 2Q\epsilon} \frac{\epsilon}{t} e^{ht}$  for all  $t \geq b$ , and thus (6.5) gives

$$(6.6) \quad e^{-2Q\epsilon} \sum_{k=0}^{\lfloor (T-b)/\epsilon \rfloor} \epsilon \frac{e^{ht_k}}{t_k} \leq \#(P(T) \setminus P(b)) \leq e^{2Q\epsilon} \sum_{k=0}^{\lceil (T-b)/\epsilon \rceil} \epsilon \frac{e^{ht_k}}{t_k}.$$

Assume that  $b$  is also chosen large enough that  $t \mapsto \frac{e^{ht}}{t}$  is nondecreasing on  $(b, \infty)$ . Then the first sum in (6.6) is an upper bound for  $\int_b^T \frac{e^{ht}}{t} dt$ , and the second sum is a lower bound for  $\int_b^{T+\epsilon} \frac{e^{ht}}{t} dt$ , so we conclude that

$$(6.7) \quad e^{-2Q\epsilon} \int_b^T \frac{e^{ht}}{t} dt \leq \#(P(T) \setminus P(b)) \leq e^{2Q\epsilon} \int_b^{T+\epsilon} \frac{e^{ht}}{t} dt.$$

Integrating by parts gives

$$(6.8) \quad \int_b^T \frac{e^{ht}}{t} dt = \frac{e^{ht}}{ht} \Big|_b^T + \int_b^T \frac{e^{ht}}{ht^2} dt \geq \frac{e^{hT}}{hT} - \frac{e^{hb}}{hb},$$

and similarly

$$\int_b^{T+\epsilon} \frac{e^{ht}}{t} dt = \frac{e^{ht}}{ht} \Big|_b^{T+\epsilon} + \int_b^{T+\epsilon} \frac{e^{ht}}{ht^2} dt \leq \frac{e^{h(T+\epsilon)}}{h(T+\epsilon)} + \frac{1}{hb} \int_b^{T+\epsilon} \frac{e^{ht}}{t} dt,$$

which yields

$$\left(1 - \frac{1}{hb}\right) \int_b^{T+\epsilon} \frac{e^{ht}}{t} dt \leq e^{h\epsilon} \frac{e^{hT}}{hT}.$$

Choosing  $b$  sufficiently large that  $1 - \frac{1}{hb} \geq e^{-h\epsilon}$ , we can combine this with (6.7) and (6.8) to obtain

$$e^{-2Q\epsilon} \left( \frac{e^{hT}}{hT} - \frac{e^{hb}}{hb} \right) \leq \#P(b, T) \leq e^{2Q\epsilon} e^{2h\epsilon} \frac{e^{hT}}{hT}.$$

Sending  $T \rightarrow \infty$  gives

$$\#P(T) \sim \#(P(T) \setminus P(b)) \sim e^{\pm 2(Q+h)\epsilon} \frac{e^{hT}}{hT}.$$

Since  $\epsilon > 0$  can be arbitrarily small, this implies (1.1) and completes the proof.

## APPENDIX A. PROOFS OF BASIC GEOMETRIC RESULTS

*Proof of Lemma 2.16.* Let  $\{\underline{c}_s: \mathbb{R}/\ell_s\mathbb{Z} \rightarrow M\}_{s \in [0,1]}$  be a homotopy between  $\underline{c}_0$  and  $\underline{c}_1$ . Lifting to a homotopy  $\{c_s: \mathbb{R} \rightarrow X\}_{s \in [0,1]}$  we observe that for each  $s \in [0,1]$  there is  $\gamma_s \in \Gamma$  such that  $c_s(t + \ell_s) = \gamma_s c_s(t)$  for all  $t \in \mathbb{R}$ . Moreover,  $s \mapsto \gamma_s$  is continuous, hence constant, so  $\gamma_1 = \gamma_0$ , which proves the lemma.  $\square$

*Proof of Lemma 2.17.* Since  $\gamma$  acts isometrically, we have

$$d(c_0(0), c_1(0)) = d(\gamma^n c_0(0), \gamma^n c_1(0)) = d(c_0(n\ell_0), c_1(n\ell_1))$$

for all  $n \in \mathbb{N}$ , and the triangle inequality gives

$$\begin{aligned} n|\ell_1 - \ell_0| &= d(c_1(n\ell_1), c_1(n\ell_0)) \\ &\leq d(c_1(n\ell_1), c_0(n\ell_0)) + d(c_0(n\ell_0), c_1(n\ell_0)) \\ &\leq d(c_0(0), c_1(0)) + d(c_0(n\ell_0), c_1(n\ell_0)). \end{aligned}$$

Lemma 2.10 gives  $c_1(\infty) = c_0(\infty)$ , so  $\sup_n d(c_0(n\ell_0), c_1(n\ell_0)) < \infty$ , and we conclude that  $\sup_n n|\ell_1 - \ell_0| < \infty$ , which implies  $\ell_1 = \ell_0$ . Applying this result to the geodesic shows that the common value is  $|\gamma|$ .

To prove that  $\underline{c}_0$  and  $\underline{c}_1$  lie in the same free homotopy class, let  $s \mapsto c_s(0)$  be any path from  $c_0(0)$  to  $c_1(0)$ , and define  $c_s(|\gamma|) := \gamma c_s(0)$ . This defines  $c_s(t)$  as a continuous function of  $(s, t)$  on the boundary of  $[0, 1] \times [0, |\gamma|]$ . Since  $X$  is simply connected this extends to a continuous map on all of  $[0, 1] \times [0, |\gamma|]$ , and then to  $[0, 1] \times \mathbb{R}$  by defining  $c_s(t \pm |\gamma|) := \gamma^{\pm 1} c_s(t)$ ; this gives the desired homotopy.  $\square$

*Proof of Lemma 3.8.* By continuity of  $H$ , it suffices to show that  $H^{-1}(\mathbf{P} \times \mathbf{F} \times \{0\})$  is bounded. Let  $H_0$  be the Hopf map for the background metric  $g_0$ . It is a homeomorphism, so  $H_0^{-1}(\mathbf{P} \times \mathbf{F} \times \{0\})$  is compact, hence bounded. Using the Morse lemma (see in particular [CKW19, Lemma 2.4]),  $H_0^{-1} \circ H: SX \rightarrow SX$  has the property that there is  $R > 0$  such that  $d(H_0^{-1} \circ H(v), v) \leq R$  for all  $v \in SX$ , and  $H^{-1}(\mathbf{P} \times \mathbf{F} \times \{0\})$  is contained in the  $R$ -ball around the bounded set  $H_0^{-1}(\mathbf{P} \times \mathbf{F} \times \{0\})$ .  $\square$

*Proof of Lemma 3.9.* Assume that this is not the case. Then there exist sequences  $\theta_n \searrow 0$  and  $v_n, w_n \in H^{-1}(\mathbf{P}_{\theta_n} \times \mathbf{F}_{\theta_n} \times \{0\})$  such that  $d(\pi v_n, \pi w_n) \geq \frac{\epsilon}{2}$ . By Lemma 3.8,  $H^{-1}(\mathbf{P}_{\theta_n} \times \mathbf{F}_{\theta_n} \times \{0\})$  is compact, so there is a subsequence  $n_k \rightarrow \infty$  such that  $v_{n_k} \rightarrow v$  and  $w_{n_k} \rightarrow w$ , with  $d(v, w) \geq \frac{\epsilon}{2}$  and  $v, w \in H^{-1}(\mathbf{P}_{\theta_n} \cap \mathbf{F}_{\theta_n} \times \{0\})$  for every  $n$ . This implies that  $v \neq w$  and that  $E(v) = E(w) = (v_0^-, v_0^+)$ , contradicting the assumption that  $v_0 \in \mathcal{E}$ .  $\square$

Lemma 3.11 is an immediate consequence of the following general result, which we prove using Corollary 3.4.

**Lemma A.1.** *Let  $M$  be a closed Riemannian manifold without conjugate points that satisfies the uniform visibility condition and admits a background metric of negative curvature, and let  $X$  be its universal*

cover. Fix  $p \in X$ . Suppose  $A \subset S_p X$  is closed and  $B \subset X$  is such that  $A^+ := \{v^+ : v \in A\}$  and  $B^\infty := \{\lim_n q_n \in \partial X : q_n \in B\}$  are disjoint subsets of  $\partial X$ . Then the family of functions  $A \rightarrow \mathbb{R}$  indexed by  $B$  and given by  $v \mapsto b_v(q)$  are equicontinuous: for every  $\epsilon > 0$  there exists  $\delta > 0$  such that if  $\angle_p(v, w) < \delta$ , then  $|b_v(q) - b_w(q)| < \epsilon$  for every  $q \in B$ .

*Proof.* We start by defining  $b_x(q, p)$  and  $\beta_p(x, y)$  when  $x, y \in X$ , not just when  $x, y \in \partial X$ . Given  $x, q, p \in X$ , let

$$(A.1) \quad b_x(q, p) := d(q, x) - d(p, x).$$

This represents how much longer it takes to get to  $x$  if you start at  $q$ , compared to starting at  $p$ . When  $x \in \partial X$ , recall that  $b_x(q, p)$  is the Busemann function defined by Definition 3.2, and if a sequence  $x_n \in X$  converges to  $x \in \partial X$ , then  $b_{x_n}(q, p) \rightarrow b_x(q, p)$ ; see [CKW19, Corollary 2.18]. Thus by Corollary 3.4,  $(x, q, p) \mapsto b_x(q, p)$  is continuous on  $\bar{X} \times X \times X$ .

Now given  $x, q, p \in X$ , let

$$(A.2) \quad \beta_p(x, y) = d(x, p) + d(y, p) - d(x, y).$$

This represents the extra distance it takes to travel via  $p$  when going from  $x$  to  $y$ . Given any point  $q$  on the geodesic from  $x$  to  $y$ , we have  $d(x, y) = d(x, q) + d(q, y)$  and thus

$$(A.3) \quad \begin{aligned} \beta_p(x, y) &= d(x, p) + d(y, p) - d(x, q) - d(y, q) \\ &= -(b_x(q, p) + b_y(q, p)). \end{aligned}$$

Comparing this to (3.3), we see that  $(p, x, y) \mapsto \beta_p(x, y)$  is a continuous function on  $X \times (\bar{X} \times \bar{X} \setminus \{(\xi, \xi) : \xi \in \partial X\})$ . (Recall that if  $x \in \partial X$  then the value of  $\beta_p(x, x)$  would be infinite.) Moreover, given  $p, q \in X$  and  $\xi \in \partial X$ , we have

$$\beta_p(q, \xi) = -b_\xi(q, p) - b_q(q, p) = -b_\xi(q, p) + d(q, p).$$

Now in the setting of Lemma A.1, we have fixed  $p \in X$  and see that for each  $q \in B$  and  $v \in A$  we have

$$(A.4) \quad b_v(q) = b_{v^+}(q, p) = d(q, p) - \beta_p(q, v^+).$$

Let  $\bar{B}$  be the set of all limit points of  $B$  in  $\bar{X}$ , and observe that  $A^+ \times \bar{B}$  is a compact set on which  $\beta_p$  is continuous; this is where we use the assumption that  $A^+ \cap B^\infty = \emptyset$ . Thus  $\beta_p$  is uniformly continuous on this set, and for every  $\epsilon > 0$  there is  $\delta > 0$  such that if  $\angle_p(v, w) < \delta$ , then  $|\beta_p(q, v^+) - \beta_p(q, w^+)| < \delta$ , in which case (A.4) gives  $|b_v(q) - b_w(q)| < \epsilon$ , which completes the proof of Lemma A.1.  $\square$

## REFERENCES

- [Bab02] Martine Babillot, *On the mixing property for hyperbolic systems*, Israel J. Math. **129** (2002), 61–76. MR 1910932

- [BC64] Richard L. Bishop and Richard J. Crittenden, *Geometry of manifolds*, Pure and Applied Mathematics, Vol. XV, Academic Press, New York-London, 1964. MR 0169148
- [BCFT18] K. Burns, V. Climenhaga, T. Fisher, and D. J. Thompson, *Unique equilibrium states for geodesic flows in nonpositive curvature*, *Geom. Funct. Anal.* **28** (2018), no. 5, 1209–1259. MR 3856792
- [BG89] Keith Burns and Marlies Gerber, *Real analytic Bernoulli geodesic flows on  $S^2$* , *Ergodic Theory Dynam. Systems* **9** (1989), no. 1, 27–45. MR 991488
- [Bow72a] Rufus Bowen, *The equidistribution of closed geodesics*, *Amer. J. Math.* **94** (1972), 413–423. MR 315742
- [Bow72b] ———, *Periodic orbits for hyperbolic flows*, *Amer. J. Math.* **94** (1972), 1–30. MR 298700
- [Bow74] ———, *Maximizing entropy for a hyperbolic flow*, *Math. Systems Theory* **7** (1974), no. 4, 300–303. MR 0385928
- [Bus92] Peter Buser, *Geometry and spectra of compact Riemann surfaces*, *Progress in Mathematics*, vol. 106, Birkhäuser Boston, Inc., Boston, MA, 1992. MR 1183224
- [CK02] M. Coornaert and G. Knieper, *Growth of conjugacy classes in Gromov hyperbolic groups*, *Geom. Funct. Anal.* **12** (2002), no. 3, 464–478. MR 1924369
- [CKW19] Vaughn Climenhaga, Gerhard Knieper, and Khadim War, *Uniqueness of the measure of maximal entropy for geodesic flows on certain manifolds without conjugate points*, 2019, arXiv:1903.09831.
- [Coo93] Michel Coornaert, *Mesures de Patterson-Sullivan sur le bord d'un espace hyperbolique au sens de Gromov*, *Pacific J. Math.* **159** (1993), no. 2, 241–270. MR 1214072
- [dC92] Manfredo Perdigão do Carmo, *Riemannian geometry*, *Mathematics: Theory & Applications*, Birkhäuser Boston, Inc., Boston, MA, 1992, Translated from the second Portuguese edition by Francis Flaherty. MR 1138207
- [Don88] Victor J. Donnay, *Geodesic flow on the two-sphere. I. Positive measure entropy*, *Ergodic Theory Dynam. Systems* **8** (1988), no. 4, 531–553. MR 980796
- [Ebe72] Patrick Eberlein, *Geodesic flow in certain manifolds without conjugate points*, *Trans. Amer. Math. Soc.* **167** (1972), 151–170. MR 295387
- [EO73] P. Eberlein and B. O'Neill, *Visibility manifolds*, *Pacific J. Math.* **46** (1973), 45–109. MR 336648
- [Esc77] Jost-Hinrich Eschenburg, *Horospheres and the stable part of the geodesic flow*, *Math. Z.* **153** (1977), no. 3, 237–251. MR 440605
- [FM82] A. Freire and R. Mañé, *On the entropy of the geodesic flow in manifolds without conjugate points*, *Invent. Math.* **69** (1982), no. 3, 375–392. MR 679763
- [Gan77] Ramesh Gangolli, *The length spectra of some compact manifolds of negative curvature*, *J. Differential Geometry* **12** (1977), no. 3, 403–424. MR 650997
- [Gre56] L. W. Green, *Geodesic instability*, *Proc. Amer. Math. Soc.* **7** (1956), 438–448. MR 79804
- [Gul75] Robert Gulliver, *On the variety of manifolds without conjugate points*, *Trans. Amer. Math. Soc.* **210** (1975), 185–201. MR 383294

- [Gun02] Roland Gunesch, *Precise asymptotics for periodic orbits of the geodesic flow in nonpositive curvature*, ProQuest LLC, Ann Arbor, MI, 2002, Thesis (Ph.D.)—The Pennsylvania State University. MR 2706087
- [Hej76] Dennis A. Hejhal, *The Selberg trace formula for  $\mathrm{PSL}(2, R)$ . Vol. I*, Lecture Notes in Mathematics, Vol. 548, Springer-Verlag, Berlin-New York, 1976. MR 0439755
- [Hub59] Heinz Huber, *Zur analytischen Theorie hyperbolischen Raumformen und Bewegungsgruppen*, Math. Ann. **138** (1959), 1–26. MR 109212
- [Kai90] Vadim A. Kaimanovich, *Invariant measures of the geodesic flow and measures at infinity on negatively curved manifolds*, vol. 53, 1990, Hyperbolic behaviour of dynamical systems (Paris, 1990), pp. 361–393. MR 1096098
- [Kat82] A. Katok, *Entropy and closed geodesics*, Ergodic Theory Dynam. Systems **2** (1982), no. 3-4, 339–365 (1983). MR 721728
- [KH95] Anatole Katok and Boris Hasselblatt, *Introduction to the modern theory of dynamical systems*, Encyclopedia of Mathematics and its Applications, vol. 54, Cambridge University Press, Cambridge, 1995, With a supplementary chapter by Katok and Leonardo Mendoza. MR 1326374
- [Kni83] Gerhard Knieper, *Das Wachstum der Äquivalenzklassen geschlossener Geodätischer in kompakten Mannigfaltigkeiten*, Arch. Math. (Basel) **40** (1983), no. 6, 559–568. MR 710022
- [Kni86] ———, *Mannigfaltigkeiten ohne konjugierte Punkte*, Bonner Mathematische Schriften [Bonn Mathematical Publications], vol. 168, Universität Bonn, Mathematisches Institut, Bonn, 1986, Dissertation, Rheinische Friedrich-Wilhelms-Universität, Bonn, 1985. MR 851010
- [Kni97] G. Knieper, *On the asymptotic geometry of nonpositively curved manifolds*, Geom. Funct. Anal. **7** (1997), no. 4, 755–782. MR 1465601
- [Kni98] Gerhard Knieper, *The uniqueness of the measure of maximal entropy for geodesic flows on rank 1 manifolds*, Ann. of Math. (2) **148** (1998), no. 1, 291–314. MR 1652924
- [Kni02] ———, *Hyperbolic dynamics and Riemannian geometry*, Handbook of dynamical systems, Vol. 1A, North-Holland, Amsterdam, 2002, pp. 453–545. MR 1928523
- [LS19] Yuri Lima and Omri M. Sarig, *Symbolic dynamics for three-dimensional flows with positive topological entropy*, J. Eur. Math. Soc. (JEMS) **21** (2019), no. 1, 199–256. MR 3880208
- [Man79] Anthony Manning, *Topological entropy for geodesic flows*, Ann. of Math. (2) **110** (1979), no. 3, 567–573. MR 554385
- [Mar69] G. A. Margulis, *Certain applications of ergodic theory to the investigation of manifolds of negative curvature*, Funkcional. Anal. i Priložen. **3** (1969), no. 4, 89–90. MR 0257933
- [Mar04] Grigoriy A. Margulis, *On some aspects of the theory of Anosov systems*, Springer Monographs in Mathematics, Springer-Verlag, Berlin, 2004, With a survey by Richard Sharp: Periodic orbits of hyperbolic flows, Translated from the Russian by Valentina Vladimirovna Szulikowska. MR 2035655
- [McK72] H. P. McKean, *Selberg’s trace formula as applied to a compact Riemann surface*, Communications on Pure and Applied Mathematics **25** (1972), no. 3, 225–246.
- [Osi10] Denis Osin, *Small cancellations over relatively hyperbolic groups and embedding theorems*, Ann. of Math. (2) **172** (2010), no. 1, 1–39. MR 2680416

- [Pat76] S. J. Patterson, *The limit set of a Fuchsian group*, Acta Math. **136** (1976), no. 3-4, 241–273. MR 0450547
- [Pes77] Ja. B. Pesin, *Geodesic flows in closed Riemannian manifolds without focal points*, Izv. Akad. Nauk SSSR Ser. Mat. **41** (1977), no. 6, 1252–1288, 1447. MR 0488169
- [PP83] William Parry and Mark Pollicott, *An analogue of the prime number theorem for closed orbits of Axiom A flows*, Ann. of Math. (2) **118** (1983), no. 3, 573–591. MR 727704
- [Ric19] Russell Ricks, *Counting closed geodesics in a compact rank one locally cat(0) space*, 2019, arXiv:1903.07635v1.
- [Rug07] Rafael O. Ruggiero, *Dynamics and global geometry of manifolds without conjugate points*, Ensaios Matemáticos [Mathematical Surveys], vol. 12, Sociedade Brasileira de Matemática, Rio de Janeiro, 2007. MR 2304843
- [Sak96] Takashi Sakai, *Riemannian geometry*, Translations of Mathematical Monographs, vol. 149, American Mathematical Society, Providence, RI, 1996, Translated from the 1992 Japanese original by the author. MR 1390760
- [Sel56] A. Selberg, *Harmonic analysis and discontinuous groups in weakly symmetric Riemannian spaces with applications to Dirichlet series*, J. Indian Math. Soc. (N.S.) **20** (1956), 47–87. MR 88511
- [Sin66] Ja. G. Sinaĭ, *Asymptotic behavior of closed geodesics on compact manifolds with negative curvature*, Izv. Akad. Nauk SSSR Ser. Mat. **30** (1966), 1275–1296. MR 0229178
- [Sul79] Dennis Sullivan, *The density at infinity of a discrete group of hyperbolic motions*, Inst. Hautes Études Sci. Publ. Math. (1979), no. 50, 171–202. MR 556586
- [Tol84] Charles Hansen Toll, *A multiplicative asymptotic for the prime geodesic theorem*, ProQuest LLC, Ann Arbor, MI, 1984, Thesis (Ph.D.)–University of Maryland, College Park. MR 2634161
- [Wal82] Peter Walters, *An introduction to ergodic theory*, Graduate Texts in Mathematics, vol. 79, Springer-Verlag, New York-Berlin, 1982. MR 648108
- [Wea14] Bryce Weaver, *Growth rate of periodic orbits for geodesic flows over surfaces with radially symmetric focusing caps*, J. Mod. Dyn. **8** (2014), no. 2, 139–176. MR 3277199

DEPT. OF MATHEMATICS, UNIVERSITY OF HOUSTON, HOUSTON, TX 77204

DEPT. OF MATHEMATICS, RUHR UNIVERSITY BOCHUM, 44780 BOCHUM, GERMANY

IMPA, ESTRADA DONA CASTORINA 110, RIO DE JANEIRO, BRAZIL

*Email address:* `climinha@math.uh.edu`

*Email address:* `gerhard.knieper@rub.de`

*Email address:* `khadim@impa.br`