CLOSED GEODESICS AND HOLONOMIES FOR KLEINIAN MANIFOLDS

GREGORY MARGULIS, AMIR MOHAMMADI, AND HEE OH

Abstract. For a rank one Lie group G and a Zariski dense and geometrically finite subgroup Γ of G, we establish the joint equidistribution of closed geodesics and their holonomy classes for the associated locally symmetric space. Our result is given in a quantitative form for geometrically finite real hyperbolic manifolds whose critical exponents are big enough. In the case when $G = \mathrm{PSL}_2(\mathbb{C})$, our results imply the equidistribution of eigenvalues of elements of Γ in the complex plane.

When Γ is a lattice, the equidistribution of holonomies was proved by Sarnak and Wakayama in 1999 using the Selberg trace formula.

1. Introduction

A rank one locally symmetric space X is of the form $\Gamma\backslash G/K$ where G is a connected simple linear Lie group of real rank one, K is a maximal compact subgroup of G and Γ is a torsion-free discrete subgroup of G. Let $o := [K] \in G/K$ and choose a unit tangent vector v_o at o. Let M denote the subgroup of G which stabilizes v_o. The unit tangent bundle $\mathrm{T}^1(X)$ of X can be identified with $\Gamma\backslash G/M$. Each closed geodesic C on $\mathrm{T}^1(X)$ gives rise to the holonomy conjugacy class h_C in M which is obtained by parallel transport about C.

Our aim in this paper is to establish the equidistribution of holonomies about closed geodesics C with length $\ell(C)$ going to infinity, when Γ is geometrically finite and Zariski dense in G. We will indeed prove a stronger joint equidistribution theorem for closed geodesics and their holonomy classes. A discrete subgroup Γ is called *geometrically finite* if the unit neighborhood of its convex core in X is of finite Riemannian volume (cf. [6]). Lattices are clearly geometrically finite, but there is also a big class of discrete subgroups of infinite co-volume which are geometrically finite. For instance, if G/K is the real hyperbolic space $\mathbb{H}^n$ and $G = \mathrm{SO}(n,1)^\circ$ is the group of its

---

2010 *Mathematics Subject Classification.* Primary 11N45, 37F35, 22E40; Secondary 37A17, 20F67.

Key words and phrases. Geometrically finite groups, Closed geodesics, Holonomy, Bowen-Margulis-Sullivan measure.

Margulis was supported in part by NSF Grant #1265695.

Mohammadi was supported in part by NSF Grant #1200388.

Oh was supported in part by NSF Grant #1068094.

orientation preserving isometries, any discrete group Γ admitting a finite-sided convex fundamental domain is geometrically finite. The fundamental group of a finite volume hyperbolic manifold with non-empty totally geodesic boundary is also known to be geometrically finite. We denote by $\delta = \delta_\Gamma$ the critical exponent of Γ. It is well-known that $\delta > 0$ if Γ is non-elementary.

In this paper, a closed geodesic in $\mathrm{T}^1(X)$ is always meant to be a *primitive* closed geodesic, unless mentioned otherwise. For $T > 0$, we set

$$\mathcal{G}_\Gamma(T) := \{C : C \text{ is a closed geodesic in } \mathrm{T}^1(X), \ell(C) \le T\}.$$

The following theorem follows from a stronger joint equidistribution theorem 5.1.

Theorem 1.1. *Let Γ be geometrically finite and Zariski dense in G. Then for any continuous class function φ on M,*

$$\sum_{C\in\mathcal{G}_\Gamma(T)} \varphi(h_C) \sim \frac{e^{\delta T}}{\delta T}\int_M \varphi\, dm \quad \text{as } T\to\infty$$

where dm is the Haar probability measure on M.

The asymptotic of $\#\mathcal{G}_\Gamma(T)$ was well-known, due to Margulis [11] for X compact, to Gangolli and Warner [8] for X noncompact but of finite volume, and to Roblin [18] for X geometrically finite:

$$\#\mathcal{G}_\Gamma(T) \sim \frac{e^{\delta T}}{\delta T}.$$

We do not rely on this result in our proof of Theorem 1.1.

If we define $\mathcal{G}_\Gamma^\dagger(T)$ to be the set of all primitive and non-primitive closed geodesics of length at most T, then it is easy to see that

$$\#\mathcal{G}_\Gamma^\dagger(T) = \#\mathcal{G}_\Gamma(T) + O(T)\cdot \#\mathcal{G}_\Gamma(T/2).$$

Therefore Theorem 1.1 remains the same if we replace $\mathcal{G}_\Gamma(T)$ by $\mathcal{G}_\Gamma^\dagger(T)$. It is worth mentioning that if one considers all geodesics, then it follows from the work of Prasad and Rapinchuk [17] that the set of all holonomy classes about closed geodesics in $\mathrm{T}^1(X)$ is dense in the space of all conjugacy classes of M.

When $G = \mathrm{SO}(n,1)^\circ$ and Γ is a co-compact lattice, Theorem 1.1 was known due to Parry and Pollicott [15], who showed that the topological mixing of the frame flow on a compact manifold implies the equidistribution of holonomies. When Γ is a lattice in a general rank one group G, Theorem 1.1 was proved by Sarnak and Wakayama [20]; their method is based on the Selberg trace formula and produces an error term. Therefore Theorem 1.1 is new only when Γ is of infinite co-volume in G. However our approach gives a more direct dynamical proof of Theorem 1.1 even in the lattice case.

Theorem 1.2. *Let Γ be a geometrically finite subgroup of $G = \mathrm{SO}(n,1)^\circ$ with $n \ge 3$. We suppose that $\delta > \max\{n-2, (n-2+\kappa)/2\}$ where κ is the* *maximum rank of all parabolic fixed points of* Γ. *Then there exists* $\eta > 0$ *such that for any smooth class function* φ *on* M,

$$\sum_{C \in \mathcal{G}_\Gamma(T)} \varphi(h_C) = \frac{e^{\delta T}}{\delta T} \int_M \varphi \, dm + O(e^{(\delta - \eta)T}) \quad \text{as } T \to \infty$$

where the implied constant depends only on the Sobolev norm of φ.

Theorem 1.2 gives a quantitative counting result for closed geodesics of length at most T. This was known when Γ is a lattice by the work of Selberg, and Gangolli-Warner [8] by the trace formula approach, or when Γ is a convex co-compact subgroup of $\mathrm{SO}(2,1)$ by Naud [13] by the symbolic dynamics approach. We remark that Theorem 1.2 can be extended to geometrically finite groups in other rank one Lie groups for which Theorem 4.4 holds; this will be evident from our proof.

As is well-known, the set of closed geodesics in $\mathrm{T}^1(X)$ is in one-to-one correspondence with the set of conjugacy classes of primitive hyperbolic elements of Γ. If $A = \{a_t\}$ denotes the one parameter subgroup whose right translation action on $\Gamma\backslash G/M$ corresponds to the geodesic flow on $\mathrm{T}^1(X)$, then any hyperbolic element $g \in G$ is conjugate to $a_g m_g$ with $a_g \in A^+ := \{a_t : t > 0\}$ and $m_g \in M$. Moreover a_g is uniquely determined, and m_g is uniquely determined up to a conjugation in M. Denote by $[\gamma]$ the conjugacy class of γ in Γ and by $[\Gamma_{ph}]$ the set of all conjugacy classes of *primitive hyperbolic* elements of Γ. Given a closed geodesic C in $\mathrm{T}^1(X)$, if $[\gamma] \in [\Gamma_{ph}]$ is the corresponding conjugacy class, then the holonomy class h_C is precisely the conjugacy class $[m_\gamma]$. Therefore Theorem 1.1 can also be interpreted as the equidistribution of $[m_\gamma]$'s among primitive hyperbolic conjugacy classes of Γ.

For $G = \mathrm{PSL}_2(\mathbb{C})$, Theorem 1.1 implies the equidistribution of eigenvalues of Γ. If we denote by λ_γ and λ_γ^{-1} the eigenvalues of $\gamma \in \Gamma$ (up to sign) so that $|\lambda_\gamma| \geq 1$, then γ is hyperbolic if and only if $|\lambda_\gamma| > 1$.

The aforementioned result of Prasad and Rapinchuk says that any Zariski dense subgroup Γ contains a hyperbolic element γ such that the argument of the complex number λ_γ is an irrational multiple of π [17]. We show a stronger theorem that the arguments of λ_γ's are equidistributed in all directions when Γ is geometrically finite.

Theorem 1.3. *Let* $G = \mathrm{PSL}_2(\mathbb{C})$ *and* Γ *be a geometrically finite and Zariski dense subgroup of* G.

For any $0 < \theta_1 < \theta_2 < \pi$, *we have*

$$\#\{[\gamma] \in [\Gamma_{ph}] : |\lambda_\gamma| < T,\ \theta_1 < \mathrm{Arg}(\lambda_\gamma) < \theta_2\} \sim \frac{(\theta_2 - \theta_1) T^{2\delta}}{2\pi\delta \log T} \quad \text{as } T \to \infty. \tag{1.4}$$

If $\delta > 1$ *and* Γ *has no rank* 2 *cusp, or if* $\delta > 3/2$ *in general, then* (1.4) *holds with a polynomial error term* $O(T^{2\delta - \varepsilon_0})$ *for some* $\varepsilon_0 > 0$.

For a hyperbolic element $\gamma \in \Gamma$, the length of the corresponding geodesic is $2\log|\lambda_\gamma|$ and the argument of λ_γ is precisely the holonomy associated to γ. Hence Theorem 1.3 is a special case of Theorems 1.1 and 1.2.

In the case when Γ is contained in an arithmetic subgroup of $\mathrm{PSL}_2(\mathbb{C})$, the polynomial error term can be taken to be uniform over all congruence subgroups of Γ; this follows from our approach based on the work of Bourgain, Gamburd and Sarnak [2] and of Mohammadi and Oh [12]. In the case when $\Gamma \subset \mathrm{PSL}_2(\mathcal{O}_D)$ where $\mathcal{O}_D$ is the ring of integers of an imaginary quadratic extension $\mathbb{Q}(\sqrt{-D})$ of $\mathbb{Q}$, the eigenvalues of Γ are fundamental units of $\mathcal{O}_D$ (cf. [19]), in which case Theorem 1.3 also bears an arithmetic application on the distribution of such fundamental units arising from Γ.

In proving Theorem 1.1, we consider the following measure μ_T on the product space $\mathrm{T}^1(X) \times M^{\mathrm{c}}$ where M^{c} denotes the space of conjugacy classes of M: for $f \in C(\mathrm{T}^1(X))$ and $\xi \in C(M^{\mathrm{c}})$, set

$$\eta_T(f \otimes \xi) := \sum_{C \in \mathcal{G}_\Gamma(T)} \mathcal{D}_C(f)\xi(h_C) \tag{1.5}$$

where $\mathcal{D}_C$ denotes the length measure on the geodesic C, normalized to be a probability measure. Theorem 1.1 follows if we show that for any *bounded* continuous function f and a continuous function ξ,

$$\eta_T(f \otimes \xi) \sim \frac{e^{\delta T} \cdot m^{\mathrm{BMS}}(f) \cdot \int_M \xi dm}{\delta \cdot |m^{\mathrm{BMS}}| \cdot T} \quad \text{as } T \to \infty \tag{1.6}$$

where m^{BMS} is the Bowen-Margulis-Sullivan measure on $\mathrm{T}^1(X)$.

We will deduce (1.6) from the following:

$$\mu_T(f \otimes \xi) \sim \frac{e^{\delta T} \cdot m^{\mathrm{BMS}}(f) \cdot \int_M \xi dm}{\delta \cdot |m^{\mathrm{BMS}}|} \quad \text{as } T \to \infty \tag{1.7}$$

where $\mu_T(f \otimes \xi) = \sum_{C \in \mathcal{G}_\Gamma(T)} \mathcal{L}_C(f)\xi(h_C)$ for $\mathcal{L}_C = \ell(C) \cdot \mathcal{D}_C$ (the length measure on C).

Let N^+ and N^- denote the expanding and contracting horospherical subgroups of G with respect to A, respectively. In studying (1.7), the following ε-flow boxes play an important role: for $g_0 \in G$, set

$$\mathfrak{B}(g_0, \varepsilon) = g_0(N_\varepsilon^+ N^- \cap N_\varepsilon^- N^+ AM) M_\varepsilon A_\varepsilon. \tag{1.8}$$

where A_ε (resp. M_ε) is the ε-neighborhood of e in A (resp. M) and $N_\varepsilon^\pm$ denotes the ε-neighborhood of e in $N^\pm$. Let $\tilde{\mathfrak{B}}(g_0, \varepsilon)$ denote the image of $\mathfrak{B}(g_0, \varepsilon)$ under the canonical projection $G \to \Gamma\backslash G/M$. Fixing a Borel subset Ω of M which is conjugation-invariant, the main idea is to relate the restriction of μ_T to $\tilde{\mathfrak{B}}(g_0, \varepsilon) \otimes \Omega$ with the counting function of the set $\Gamma \cap \mathfrak{B}(g_0, \varepsilon) A_T^+ \Omega \mathfrak{B}(g_0, \varepsilon)^{-1}$ with $A_T^+ = \{a_t : 0 < t \le T\}$ (see Comparison lemma 5.12); we establish this relation using the effective closing lemma 3.1. We remark that for the effective closing lemma, it is quite essential to use a flow box which is precisely of the form given in (1.8). This flow box was first used in Margulis' work on counting closed geodesics [11]. The counting function of $\Gamma \cap \mathfrak{B}(g_0, \varepsilon) A_T^+ \Omega \mathfrak{B}(g_0, \varepsilon)^{-1}$ can then be understood based on the mixing result of Winter [24], which says that the A action on $L^2(\Gamma\backslash G, m^{\mathrm{BMS}})$ is mixing. An effective mixing statement for the cases mentioned in Theorem 1.2 was obtained in [12]. We also remark that if we restrict ourselves only to those f with compact support, then (1.7) holds for any discrete subgroup Γ admitting a finite BMS measure; that is, Γ need not be geometrically finite.

Acknowledgement We would like to thank Dale Winter for helpful comments on the preprint. We also thank the referee for the careful reading of our manuscript and helpful comments.

2. Preliminaries

Throughout the paper, let G be a connected simple real linear Lie group of real rank one. As is well known, G is one of the following type: $\mathrm{SO}(n,1)^\circ$, $\mathrm{SU}(n,1)$, $\mathrm{Sp}(n,1)$ $(n \geq 2)$ and F_4^{-20}, which are the groups of isometries of the hyperbolic spaces $\mathbb{H}^n_{\mathbb{R}}$, $\mathbb{H}^n_{\mathbb{C}}$, $\mathbb{H}^n_{\mathbb{H}}$, $\mathbb{H}^2_{\mathbb{O}}$ respectively. Let K be a maximal compact subgroup of G. Then $\tilde{X} := G/K$ is a symmetric space of rank one. Let $o \in \tilde{X}$ be the point which is stabilized by K. The killing form on the Lie algebra of G endows a left G-invariant metric $d_{\tilde{X}}$ on $\tilde{X}$ which we normalize so that the maximum sectional curvature is -1.

The volume entropy $D(\tilde{X})$ of $\tilde{X}$ is defined by

$$D(\tilde{X}) = \lim_{T\to\infty} \frac{\log \mathrm{Vol}(B(o,T))}{T} \tag{2.1}$$

where $B(o,T) = \{x \in \tilde{X} : d_{\tilde{X}}(o,x) \leq T\}$. It is explicitly given as follows:

$$D(\tilde{X}) = n-1, 2n, 4n+2, 22 \tag{2.2}$$

respectively for $\mathrm{SO}(n,1)^\circ$, $\mathrm{SU}(n,1)$, $\mathrm{Sp}(n,1)$ and F_4^{-20}.

We denote by $\partial_\infty(\tilde{X})$ the geometric boundary of $\tilde{X}$ and by $\mathrm{T}^1(\tilde{X})$ the unit tangent bundle of $\tilde{X}$. Fixing a vector $v_o \in \mathrm{T}^1(\tilde{X})$ based at o, $\mathrm{T}^1(\tilde{X})$ can be identified with G/M where M is the stabilizer of v_o in G. For a vector $v \in \mathrm{T}^1(\tilde{X})$, we denote by $v^+ \in \partial_\infty(\tilde{X})$ and $v^- \in \partial_\infty(\tilde{X})$ the forward and the backward end points of the geodesic determined by v. For $g \in G$, we set $g^\pm = (gv_o)^\pm$. There exists a one parameter subgroup $A = \{a_t : t \in \mathbb{R}\}$ of diagonalizable elements of G which commutes with M and whose right translation action on G/M by a_t corresponds to the geodesic flow for time t on $\mathrm{T}^1(\tilde{X})$; in fact, M is equal to the centralizer of A in K. We set

$$A^+ := \{a_t : t > 0\} \quad \text{and} \quad A_T^+ := \{a_t : 0 < t \leq T\}.$$

We denote by N^+ and N^- the expanding and contracting horospherical subgroups:

$$N^+ = \{g \in G : \lim_{t\to+\infty} a_t g a_{-t} \to e\};$$

$$N^- = \{g \in G : \lim_{t\to+\infty} a_{-t} g a_t \to e\}.$$

The stabilizer of v_o^+ and v_o^- in G are given respectively by

$$P^- := MAN^-, \quad \text{and} \quad P^+ := MAN^+.$$

Hence the orbit map $g \mapsto gv_o^+$ (resp. $g \mapsto gv_o^-$) induces a homeomorphism between G/P^- (resp. G/P^+) with $\partial_\infty(\tilde{X})$.

Let $d = d_G$ be a left G-invariant Riemannian metric on G which induces the metric $d_{\tilde{X}}$ on $\tilde{X} = G/K$. For a subset S of G and $g_0 > 0$, we set

$$S_\varepsilon(g_0) := \{s \in S : d_G(g_0, s) < \varepsilon\}$$

the intersection of the ε-ball at g_0 with S. Hence the ε-balls $G_\varepsilon(g_0)$ form a basis of open neighborhoods at g_0.

Flow box: Following Margulis [11], we will define the flow-box around $g_0 \in G$ for all small $\varepsilon > 0$. For this, we will use the following ε-neighborhoods of e in N^+, N^-, A, M.

The groups $N^\pm$ are connected unipotent groups and hence the exponential map $\exp : \mathrm{Lie}(N^\pm) \to N^\pm$ is a diffeomorphism. For $\varepsilon > 0$, we set

$$N_\varepsilon^\pm := \{n_x^\pm := \exp x \in N^\pm : \|x\| < \varepsilon\}$$

where $\|x\|$ denotes a norm on the real vector space $\mathrm{Lie}(N^\pm)$ which is M-invariant under the adjoint action of M on $\mathrm{Lie}(N^\pm)$.

For A and M, we simply put

$$A_\varepsilon = A \cap G_\varepsilon(e) = \{a_t : t \in (-\varepsilon, \varepsilon)\}, \text{ and } M_\varepsilon = M \cap G_\varepsilon(e).$$

We now define the ε-flow box $\mathfrak{B}(g_0, \varepsilon)$ at g_0 as follows:

$$\mathfrak{B}(g_0, \varepsilon) = g_0(N_\varepsilon^+ N^- \cap N_\varepsilon^- N^+ AM) M_\varepsilon A_\varepsilon. \tag{2.3}$$

For simplicity, we set $\mathfrak{B}(\varepsilon) := \mathfrak{B}(e, \varepsilon)$. The product maps $N^+ \times A \times M \times N^- \to G$ and $N^- \times A \times M \times N^+ \to G$ are diffeomorphisms onto Zariski open neighborhoods of e in G. Therefore the sets $\mathfrak{B}(g_0, \varepsilon)$, $\varepsilon > 0$ form a basis of neighborhoods of g_0 in G.

We remark that this definition of the flow box is quite essential in our proof of the effective closing lemma 3.1. We list the following properties of the flow box which we will use later:

Lemma 2.4 (Basic properties of the flow box)**.** *Let $g_0 \in G$ and $\varepsilon > 0$.*

(1) *For any $g \in \mathfrak{B}(g_0, \varepsilon)$, the set $\{t \in \mathbb{R} : ga_t \in \mathfrak{B}(g_0, \varepsilon)\}$ is of Lebesgue length 2ε;*
(2) *$\mathfrak{B}(g_0, \varepsilon)v_o^+ = g_0 N_\varepsilon^- v_o^+$ and $\mathfrak{B}(g_0, \varepsilon)v_o^- = g_0 N_\varepsilon^+ v_o^-$;*
(3) *There exists $c > 1$ such that*

$$G_{c^{-1}\varepsilon}(g_0) \subset \mathfrak{B}(g_0, \varepsilon) \subset G_{c\varepsilon}(g_0); \tag{2.5}$$

here c is independent of $g_0 \in G$ and all small $\varepsilon > 0$.

Considering the action of $g \in G$ on the compactification $\tilde{X} \cup \partial(\tilde{X})$, g is called *elliptic*, *parabolic*, *hyperbolic* if the set $\mathrm{Fix}(g) = \{x \in \tilde{X} \cup \partial(\tilde{X}) : g(x) = x\}$ of fixed points by g is contained in $\tilde{X}$, is a singleton on $\partial(\tilde{X})$, and consists of two distinct points on $\partial(\tilde{X})$ respectively. Any element g in a rank one Lie group is one of these three types.

Equivalently, $g \in G$ is *elliptic* if g is conjugate to an element of K, and *parabolic* if g is conjugate to an element of $MN^+ - M$, and *hyperbolic* if g is conjugate to an element of $A^+M - M$.

Lemma 2.6. *Suppose that for some $h \in G$, $ha_1m_1h^{-1} = a_2m_2$ with $a_1, a_2 \in A^+$ and $m_1, m_2 \in M$. Then $a_1 = a_2$, $m_1 = mm_2m^{-1}$ for some $m \in M$ and $h \in AM$.*

Proof. For $g \in G$, define

$$N^{\pm}(g) = \{q \in G : g^{\ell} q g^{-\ell} \to e \text{ as } \ell \to \pm\infty\}.$$

Putting $g_i = a_i m_i \in A^+M$ for $i = 1, 2$, we have $N^{\pm}(g_i) = N^{\pm}$. On the other hand, since $g_2 = hg_1h^{-1}$, the above definition implies $N^{\pm}(g_2) = hN^{\pm}(g_1)h^{-1}$. Hence h belongs to the common normalizer of $N^{\pm}$, which is equal to $P^+ \cap P^- = AM$. Therefore $h = am \in AM$. It now follows that $ha_1m_1h^{-1} = a_1(mm_1m^{-1}) = a_2m_2$. Hence $a_2^{-1}a_1 \in A \cap M = \{e\}$; so $a_1 = a_2$, as well as $m_2 = mm_1m^{-1}$. $\square$

As an immediate corollary, we have:

Corollary 2.7. *If a hyperbolic element $g \in G$ is of the form:*

$$g = h_g a_g m_g h_g^{-1} \tag{2.8}$$

with $a_g m_g \in A^+M$, then a_g is uniquely determined, $m_g \in M$ is determined unique up to conjugation and $R_g := h_g A v_o$ is independent of the choice of h_g.

The geodesic $R_g := h_g A v_o \subset \tilde{X}$ is called *the oriented axis* of g: g preserves R_g, and acts as a translation by $T := d(a_g, e)$.

Let Γ be a torsion-free and non-elementary discrete subgroup of G. A closed geodesic C of length $T > 0$ on $\mathrm{T}^1(X) = \Gamma\backslash G/M$ is a compact set of the form $\Gamma\backslash\Gamma gAM/M$ for some $g \in G$ such that $gAMg^{-1} \cap \Gamma$ is generated by a hyperbolic element $\gamma = g a_\gamma m_\gamma g^{-1}$ with $T = d(a_\gamma, e)$. The conjugacy class $[m_\gamma]$ in M is called the holonomy class attached to C. Note that if we have

$$\Gamma\backslash\Gamma g m_0 a_T = \Gamma\backslash\Gamma g m_0 m$$

for some $m_0, m \in M$, then $[m] = [m_\gamma]$. Geometrically, $\Gamma\backslash\Gamma g m_0$ is a frame which contains the tangent vector $\Gamma\backslash\Gamma gM$, and the element m measures the extent to which parallel transport around the closed geodesic $\Gamma g m_0 a_T$ differs from the original frame $\Gamma\backslash\Gamma g m_0$. If we choose a different base point m_1 from m_0, then m changes by a conjugation; hence the holonomy class attached to C is well-defined.

FIGURE 1. Pictorial proof of Closing lemma

3. Effective closing lemma

Let Γ be a torsion-free, non-elementary and discrete subgroup of G and set $X := \Gamma\backslash G/K$, which is a rank one locally symmetric manifold whose fundamental group is isomorphic to Γ. We denote by $\pi : G \to \Gamma\backslash G$ the canonical projection map.

For two elements $h_1, h_2 \in G$, we will write $h_1 \sim_\varepsilon h_2$ if $d_G(h_1, h_2) \le \varepsilon$ and $h_1 \sim_{O(\varepsilon)} h_2$ if $d_G(h_1, h_2) \le c\varepsilon$ for some constant $c > 1$ depending only on G. For conjugacy classes $[m_1]$ and $[m_2]$ in M, we write $[m_1] \sim_\varepsilon [m_2]$ and $[m_1] \sim_{O(\varepsilon)} [m_2]$ if, respectively, $m_1 \sim_\varepsilon m_2$ and $m_1 \sim_{O(\varepsilon)} m_2$ for some representatives $m_i \in M$ of $[m_i]$.

For a subset S of G and $\varepsilon > 0$, we also use the notation $G_{O(\varepsilon)}(S)$ for the $c\varepsilon$-neighborhood of S for some $c > 1$ depending only on G, and the notation 1_S for the characteristic function of S.

For $g_0 \in G$, we will define the injectivity radius of g_0 in $\Gamma\backslash G$ to be the supremum $\varepsilon > 0$ such that the ε flow box $\mathfrak{B}(g_0, \varepsilon)$ injects to $\Gamma\backslash G$. In what follows, we will consider boxes $\mathfrak{B}(g_0, \varepsilon)$ only for those ε which are smaller than the injectivity radius of g_0, without repeatedly saying so.

In this section, we consider the situation where a long geodesic comes back to a fixed ε-box $\pi(\mathfrak{B}(g_0, \varepsilon))$, that is, there exist $g_1, g_2 \in \mathfrak{B}(g_0, \varepsilon)$ such that

$$g_1 \tilde{a}_\gamma \tilde{m}_\gamma = \gamma g_2$$

for some $\gamma \in \Gamma$ and $\tilde{a}_\gamma \tilde{m}_\gamma \in AM$ with $T := d(\tilde{a}_\gamma, e)$ sufficiently large. The so-called closing lemma for a negatively curved space (see [11, Lemma 6.2] and [18, Chapter 5]) says that there is a closed geodesic nearby; more precisely, γ is a hyperbolic element and its oriented axis R_γ is nearby the box $\pi(\mathfrak{B}(g_0, \varepsilon))$ in the space $\Gamma\backslash G/M = \mathrm{T}^1(X)$. We will need more detailed information on this situation. We will show that the oriented axis R_γ passes through $O(\varepsilon e^{-T})$-neighborhood of the box $\pi(\mathfrak{B}(g_0, \varepsilon))$ and $\tilde{a}_\gamma$ and $[\tilde{m}_\gamma]$ are $O(\varepsilon)$-close to a_γ and $[m_\gamma]$ respectively where a_γ and $[m_\gamma]$ are defined as in (2.8) for γ.

Lemma 3.1 (Effective closing lemma I)**.** *There exists $T_0 \gg 1$, depending only on G, for which the following holds: For any $g_0 \in G$ and any small*

$\varepsilon > 0$, suppose that there exist $g_1, g_2 \in \mathfrak{B}(g_0, \varepsilon)$ and $\gamma \in G$ such that

$$g_1 \tilde{a}_\gamma \tilde{m}_\gamma = \gamma g_2 \tag{3.2}$$

for some $\tilde{a}_\gamma \in A$ with $T := d(\tilde{a}_\gamma, e) \geq T_0$ and $\tilde{m}_\gamma \in M$.

Then there exists $g \in \mathfrak{B}(g_0, \varepsilon + O(\varepsilon e^{-T}))$ such that

$$\gamma = g a_\gamma m_\gamma g^{-1}.$$

Moreover, $a_\gamma \sim_{O(\varepsilon)} \tilde{a}_\gamma$, and $[m_\gamma] \sim_{O(\varepsilon)} [\tilde{m}_\gamma]$.

Proof. The proof is divided into two parts.

Step 1: We will show that for some $g_3 := g_0 h_\varepsilon \in \mathfrak{B}(g_0, \varepsilon)$,

$$g_3^{-1} \gamma g_3 = n_w^+ a'_\gamma m'_\gamma n_z^- \tag{3.3}$$

where $a'_\gamma m'_\gamma \in \tilde{a}_\gamma \tilde{m}_\gamma A_{O(\varepsilon)} M_{O(\varepsilon)}$, $n_w^+ \in N^+_{O(e^{-T}\varepsilon)}$ and $n_z^- \in N^-_{O(e^{-T}\varepsilon)}$.

To prove this claim, note that there exist $b_\varepsilon, d_\varepsilon \in \mathfrak{B}_\varepsilon(e)$ such that $g_1 = g_0 b_\varepsilon$ and $g_2 = g_0 d_\varepsilon$. Recalling the definition

$$\mathfrak{B}(\varepsilon) = (N_\varepsilon^+ N^- \cap N_\varepsilon^- N^+ AM) M_\varepsilon A_\varepsilon,$$

we may write b_ε and d_ε as follows:

$$b_\varepsilon = b_\varepsilon^+ n_x^- b_\varepsilon^0 \in N_\varepsilon^+ N^- (A_\varepsilon M_\varepsilon);$$
$$d_\varepsilon = d_\varepsilon^- n_y^+ d_\varepsilon^0 \in N_\varepsilon^- N^+ (AM).$$

By Lemma 2.4, we have $n_x^- \in N^-_{O(\varepsilon)}$, $n_y^+ \in N^+_{O(\varepsilon)}$ and $d_\varepsilon^0 \in A_{O(\varepsilon)} M_{O(\varepsilon)}$. Now the equality $g_1 \tilde{a}_\gamma \tilde{m}_\gamma = \gamma g_2$ can be rewritten as

$$g_0 b_\varepsilon^+ n_x^- \tilde{a}_\gamma^{(1)} \tilde{m}_\gamma^{(1)} = \gamma g_0 d_\varepsilon^- n_y^+ \tag{3.4}$$

where $a_\gamma^{(1)} \tilde{m}_\gamma^{(1)} := b_\varepsilon^0 \tilde{a}_\gamma \tilde{m}_\gamma (d_\varepsilon^0)^{-1} \in AM$.

By the transversality between N^- and AMN^+, we obtain a unique element $n_{x'}^- \in N^-_{O(\varepsilon)}$ satisfying that

$$b_\varepsilon^+ n_{x'}^- \in d_\varepsilon^- (N^+_{O(\varepsilon)} A_{O(\varepsilon)} M_{O(\varepsilon)}). \tag{3.5}$$

Since $b_\varepsilon^+ \in N_\varepsilon^+$ and $d_\varepsilon^- \in N_\varepsilon^-$, we have

$$h_\varepsilon := b_\varepsilon^+ n_{x'}^- \in (N_\varepsilon^+ N^- \cap N_\varepsilon^- N^+ AM) \subset \mathfrak{B}(\varepsilon)$$

and hence

$$g_0 h_\varepsilon \in \mathfrak{B}(g_0, \varepsilon).$$

Now setting $g_3 := g_0 h_\varepsilon$, we claim that (3.3) holds. By (3.5), $h_\varepsilon = d_\varepsilon^- n_v^+ a_\varepsilon m_\varepsilon \in d_\varepsilon^- (N^+_{O(\varepsilon)} A_{O(\varepsilon)} M_{O(\varepsilon)})$. Rewriting (3.4), we have

$$g_0 h_\varepsilon (n_{x'}^-)^{-1} n_x^- \tilde{a}_\gamma^{(1)} \tilde{m}_\gamma^{(1)} = \gamma g_0 h_\varepsilon m_\varepsilon^{-1} a_\varepsilon^{-1} (n_v^+)^{-1} n_y^+$$

and hence

$$\begin{aligned} g_3^{-1} \gamma g_3 &= (n_{x'}^-)^{-1} n_x^- \left(\tilde{a}_\gamma^{(1)} \tilde{m}_\gamma^{(1)} a_\varepsilon m_\varepsilon \right) \left(a_\varepsilon^{-1} m_\varepsilon^{-1} (n_y^+)^{-1} n_v^+ a_\varepsilon m_\varepsilon \right) \\ &= n_{z_1}^- \tilde{a}_\gamma^{(2)} \tilde{m}_\gamma^{(2)} n_{w_1}^+ \end{aligned}$$

where $a_\gamma^{(2)}\tilde m_\gamma^{(2)} := \tilde a_\gamma^{(1)}\tilde m_\gamma^{(1)} a_\varepsilon m_\varepsilon$, $n_{z_1}^- := (n_{x'}^-)^{-1} n_x^- \in N^-_{O(\varepsilon)}$, and $n_{w_1}^+ := a_\varepsilon^{-1} m_\varepsilon^{-1} (n_y^+)^{-1} n_v^+ a_\varepsilon m_\varepsilon \in N^+_{O(\varepsilon)}$.

We have $n_{z_2}^- := (a_\gamma^{(2)}\tilde m_\gamma^{(2)})^{-1} n_{z_1}^- \tilde a_\gamma^{(2)}\tilde m_\gamma^{(2)} \in N^-_{O(e^{-T}\varepsilon)}$ and we can write

$$n_{z_2}^- n_{w_1}^+ = n_{w_2}^+ a'_\varepsilon m'_\varepsilon n_{z_3}^- \in N^+_{O(\varepsilon)} A_{O(\varepsilon)} M_{O(\varepsilon)} N^-_{O(e^{-T}\varepsilon)}.$$

Therefore

$$g_3^{-1}\gamma g_3 = \tilde a_\gamma^{(2)}\tilde m_\gamma^{(2)} n_{z_2}^- n_{w_1}^+ = n_{w_3}^+ \tilde a'_\gamma \tilde m'_\gamma n_{z_3}^-$$

where $a'_\gamma m'_\gamma := \tilde a_\gamma^{(2)}\tilde m_\gamma^{(2)} a'_\varepsilon m'_\varepsilon$, and $n_{w_3}^+ = \tilde a_\gamma^{(2)}\tilde m_\gamma^{(2)} n_{w_2}^+ (a_\gamma^{(2)}\tilde m_\gamma^{(2)})^{-1} \in N^+_{O(e^{-T}\varepsilon)}$. This proves the claim (3.3).

Step 2: Set $g := g_3^{-1}\gamma g_3$ so that $g = n_w^+ a'_\gamma m'_\gamma n_z^-$. We claim that

$$g \in (n_x^+ n_y^-) a'_\gamma A_{O(\varepsilon)} m'_\gamma M_{O(\varepsilon)} (n_x^+ n_y^-)^{-1} \tag{3.6}$$

with $n_x^+ \in N^+_{O(\varepsilon e^{-T})}$ and $n_y^- \in N^-_{O(\varepsilon e^{-T})}$.

For n_z^- as above, for any $n_x^+ \in N_\varepsilon^+$, there exists a unique element $n^+_{\alpha(x)} \in N^+_{O(\varepsilon)}$ such that $(n_z^-) n_x^+ \in n^+_{\alpha(x)} A_\varepsilon M_\varepsilon N_\varepsilon^-$. Moreover the map $n_x^+ \mapsto n_{\alpha(x)}$ is a diffeomorphsim of N_ε^+ onto its image, which is contained in $N^+_{O(\varepsilon)}$.

Therefore the implicit function theorem implies that the map $n_x^+ \mapsto n_x^+ (a'_\gamma m'_\gamma)(n^+_{\alpha(x)})^{-1}(a'_\gamma m'_\gamma)^{-1}$ defines a diffeomorphism of N_ε^+ onto its image $N^+_{\varepsilon+O(e^{-T}\varepsilon)}$. Since $n_w^+ \in N^+_{O(e^{-T}\varepsilon)}$, if T is large enough, we can find $n_x^+ \in N_{O(e^{-T}\varepsilon)}$ such that

$$n_w^+ = n_x^+ (a'_\gamma m'_\gamma)(n^+_{\alpha(x)})^{-1}(a'_\gamma m'_\gamma)^{-1}.$$

Fixing this element n_x^+, we write $(n^+_{\alpha(x)})^{-1} n_z^- = a_\varepsilon m_\varepsilon n_u^- (n_x^+)^{-1}$ with $n_u^- \in N^-_{O(e^{-T}\varepsilon)}$, $a_\varepsilon \in A_{O(\varepsilon)}$ and $m_\varepsilon \in M_{O(\varepsilon)}$. Therefore, plugging in these,

$$\begin{aligned} g &= n_w^+ a'_\gamma m'_\gamma n_z^- \\ &= n_x^+ a'_\gamma m'_\gamma (n^+_{\alpha(x)})^{-1} n_z^- \\ &= n_x^+ (a''_\gamma m''_\gamma) n_u^- (n_x^+)^{-1} \end{aligned}$$

where $a''_\gamma = a'_\gamma a_\epsilon$ and $m''_\gamma = m'_\gamma m_\epsilon$. Since the map

$$n_y^- \mapsto (a''_\gamma m''_\gamma)^{-1} n_y^- (a''_\gamma m''_\gamma)(n_y^-)^{-1}$$

is a diffeomorphism of N_ε^- onto its image $N^-_{\varepsilon+O(e^{-T}\varepsilon)}$, for all large T, we can find $n_y^- \in N^+_{O(\varepsilon e^{-T})}$ such that

$$n_u^- = (a''_\gamma m''_\gamma)^{-1}(n_y^-)^{-1}(a''_\gamma m''_\gamma) n_y^-.$$

This yields

$$g = n_x^+ (n_y^-)^{-1} (a''_\gamma m''_\gamma) n_y^- (n_x^+)^{-1}$$

as desired.

Hence

$$\gamma = g_4 a''_\gamma m''_\gamma g_4^{-1}$$

with $g_4 := g_3 n_x^+ (n_y^-)^{-1} \in \mathfrak{B}(g_0, \varepsilon + O(\varepsilon e^{-T}))$. Therefore $a_\gamma = a''_\gamma \sim_{O(\varepsilon)} \tilde{a}_\gamma$ and $[m_\gamma] = [m''_\gamma] \sim_{O(\varepsilon)} [\tilde{m}_\gamma]$. □

Although we will only be using the above version of the closing lemma 3.1, we record the following reformulation as well, which is of more geometric flavor.

Lemma 3.7 (Effective closing lemma II)**.** *There exists $T_0 \gg 1$, depending only on G, for which the following holds: Let $g_0 \in G$ and let $\varepsilon > 0$ be smaller than the injectivity radius of g_0 in $\Gamma\backslash G$. Suppose that there exist $g_1, g_2 \in \mathfrak{B}(g_0, \varepsilon)$ and $\gamma \in \Gamma$ such that*

$$g_1 \tilde{a}_\gamma \tilde{m}_\gamma = \gamma g_2 \tag{3.8}$$

for some $\tilde{a}_\gamma \in A$ with $T := d(\tilde{a}_\gamma, e) \geq T_0$ and $\tilde{m}_\gamma \in M$. Suppose also that γ is primitive, i.e., γ cannot be written as a power of another element of Γ. Then there exists an element $g_\gamma \in \mathfrak{B}(g_0, \varepsilon + O(\varepsilon e^{-T}))$ such that

1. *the AM-orbit $\Gamma\backslash\Gamma g_\gamma AM$ is compact;*
2. *γ is a generator of the group $g_\gamma AM g_\gamma^{-1} \cap \Gamma$;*
3. *the length of the closed geodesic $C_\gamma = \Gamma\backslash\Gamma g_\gamma A(v_o)$ is $T + O(\varepsilon)$;*
4. *the holonomy class $[m_\gamma]$ associated to C_γ is within $O(\varepsilon)$-distance from $[\tilde{m}_\gamma]$.*

4. Counting results for $\Gamma \cap \mathfrak{B}(g_0, \varepsilon) A_T \Omega \mathfrak{B}(g_0, \varepsilon)^{-1}$

Let G, Γ, X, o, v_o etc be as in the previous section. Recall $A_T^+ = \{a_t : 0 < t < T\}$. Our approach of understanding the distribution of closed geodesics in $\mathrm{T}^1(X)$ passing through the flow box $\mathfrak{B}(g_0, \varepsilon)$ and with holonomy class contained in a fixed compact subset Ω of M is to interpret it as a counting problem for the set $\Gamma \cap \mathfrak{B}(g_0, \varepsilon) A_T^+ \Omega \mathfrak{B}(g_0, \varepsilon)^{-1}$ as $T \to \infty$. We will be able to approximate $\#\Gamma \cap \mathfrak{B}(g_0, \varepsilon) A_T^+ \Omega \mathfrak{B}(g_0, \varepsilon)^{-1}$ by the counting function for the intersection of Γ with a certain compact subset given in the $g_0 N^+ AMN^- g_0^{-1}$ coordinates.

In the first part of this section, we will investigate the asymptotic behavior of the following

$$\#\Gamma \cap g_0 \Xi_1 A_T^+ \Omega \Xi_2 g_0^{-1}$$

for given bounded Borel subsets $\Xi_1 \subset N^+$, $\Xi_2 \subset N^-$ and $\Omega \subset M$. In the second part, we will use this result to obtain an asymptotic formula of $\#\Gamma \cap \mathfrak{B}(g_0, \varepsilon) A_T^+ \Omega \mathfrak{B}(g_0, \varepsilon)^{-1}$.

4.1. **On the counting for $\Gamma \cap g_0 \Xi_1 A_T \Omega \Xi_2 g_0^{-1}$.** This problem can be answered under the extra assumption that Γ is Zariski dense and that the Bowen-Margulis-Sullivan measure, the BMS measure for short, on $\mathrm{T}^1(X) = \Gamma\backslash G/M$ is *finite*. The key ingredient in this case is that the M-invariant extension of the BMS measure on $\Gamma\backslash G$ is mixing for the A-action.

We begin the discussion by recalling the definition of the BMS measure. Let $\Lambda(\Gamma)$ denote the limit set of Γ, which is the set of all accumulation points in $\tilde{X} \cup \partial(\tilde{X})$ of an orbit of Γ in $\tilde{X}$. Denote by $\delta = \delta_\Gamma$ the critical exponent of Γ. Denote by $\{\nu_x : x \in \tilde{X}\}$ a Γ-invariant conformal density of dimension δ supported on the limit set $\Lambda(\Gamma)$; such a density exists by the construction given by Patterson [16]. For $\xi_1 \neq \xi_2 \in \partial(\tilde{X})$, and $x \in \tilde{X}$, we denote by $\langle \xi_1, \xi_2 \rangle_x$ the Gromov product at x. Then the visual distance on $\partial(\tilde{X})$ at x is given by

$$d_x(\xi_1, \xi_2) = e^{-\langle \xi_1, \xi_2 \rangle_x}$$

with the convention that $d_x(\xi, \xi) = 0$. The Hopf parametrization of $\mathrm{T}^1(\tilde{X})$ as $(\partial^2(\tilde{X}) - \text{Diagonal}) \times \mathbb{R}$ is given by $u \mapsto (u^+, u^-, s = \beta_{u^-}(o, u))$ where $\beta_\xi(x, y)$ denotes the Busemann function for $\xi \in \partial(\tilde{X})$, and $x, y \in \tilde{X}$. The BMS measure on $\mathrm{T}^1(\tilde{X})$ with respect to $\{\nu_x\}$ is defined as follows:

$$d\tilde{m}^{\mathrm{BMS}}(u) = \frac{d\nu_x(u^+) d\nu_x(u^-) ds}{d_x(u^+, u^-)^{2\delta}}.$$

The definition is independent of $x \in \tilde{X}$ and $\tilde{m}^{\mathrm{BMS}}$ is right A-invariant and left Γ-invariant, and hence induces a geodesic flow invariant Borel measure on $\mathrm{T}^1(X)$, which we denote by m^{BMS}. If $|m^{\mathrm{BMS}}| < \infty$, then the geodesic flow is ergodic with respect to m^{BMS}, as shown by Sullivan [21] and moreover mixing by Babillot [1].

As we are eventually interested in counting a Γ orbit in a family $\Xi_1 A_T^+ \Omega \Xi_2$ with Ω any Borel subset in M, we need to understand the mixing phenomenon for the A-action on $\Gamma\backslash G$, not only on $\Gamma\backslash G/M$. By abuse of notation, we denote by m^{BMS} the M-invariant lift of m^{BMS} to $\Gamma\backslash G$. Winter [24] showed that if Γ is Zariski dense and $|m^{\mathrm{BMS}}| < \infty$, then the A-action on $\Gamma\backslash G$ is mixing for this extension m^{BMS}; this was earlier claimed in [7] for the case of $G = \mathrm{SO}(n,1)^\circ$ and Γ geometrically finite.

In the rest of this section, we assume that

$$\Gamma \text{ is Zariski dense and } |m^{\mathrm{BMS}}| < \infty.$$

For the application of the mixing in counting problems, it is easier to use the following version on the asymptotic behavior of the matrix coefficients in Haar measure. To state this result, we need to recall the Burger-Roblin measures for the N^+ and N^- actions.

Using the homemorphism of G with $K/M \times M \times A \times N^\pm$, we define the Burger-Roblin measures $\tilde{m}^{\mathrm{BR}}$ (invariant under the N^+-action) and $\tilde{m}^{\mathrm{BR}}_*$ (invariant under the N^--action) on G as follows:

$$d\tilde{m}^{\mathrm{BR}}(kma_rn^+) = e^{-\delta r} dn^+ ds d\nu_o(kv_o^-) dm \quad \text{for } kma_rn^+ \in (K/M)MAN^+; \tag{4.1}$$

$$d\tilde{m}^{\mathrm{BR}}_*(kma_rn^-) = e^{\delta r} dn^- ds d\nu_o(kv_o^+) dm \ \text{ for } kma_rn^- \in (K/M)MAN^- \tag{4.2}$$

where dm denotes the M-invariant probability measure on M; Since M fixes v_o and hence fixes $v_o^\pm$, these measures are well-defined. The Haar measure on G is given by: for $g = a_s n^{\pm} k \in AN^{\pm}K$,

$$dg = d\tilde{m}^{\mathrm{Haar}}(a_s n^{\pm} k) = ds dn^{\pm} dk$$

where dk is the probability Haar measure on K. These measures are all left Γ-invariant and we use the notations m^{BR}, m_*^{BR}, m^{Haar} (or dg) respectively for the corresponding induced right M-invariant measures on $\Gamma\backslash G$.

The following theorem can be deduced from the mixing of m^{BMS}, as observed first by Roblin for M-invariant functions ([18], see also [14]).

Theorem 4.3. ([18], [14], [24]) *For any functions $\Psi_1, \Psi_2 \in C_c(\Gamma\backslash G)$,*

$$\lim_{t\to+\infty} e^{(D-\delta)t} \int_{\Gamma\backslash G} \Psi_1(ga_t)\Psi_2(g)dg = \frac{m^{\mathrm{BR}}(\Psi_1)\cdot m_*^{\mathrm{BR}}(\Psi_2)}{|m^{\mathrm{BMS}}|}$$

where $D = D(\tilde{X})$ is the volume entropy of $\tilde{X} = G/K$ (see (2.1) *and* (2.2)*).*

The quotient by Γ of the convex hull of $\Lambda(\Gamma)$ is called the convex core of Γ. A discrete group Γ is called *geometrically finite* if the volume of a unit neighborhood of the convex core of Γ is finite. Clearly lattices are geometrically finite. If Γ is geometrically finite, then m^{BMS} is known to be finite and the critical exponent is known to be equal to the Haudorff dimension of $\Lambda(\Gamma)$ ([22] and [4]).

We use the standard asymptotic "big-O" and "little-o" notations, where for functions $f, g : \mathbb{R}^+ \to \mathbb{R}^+$, we write $f = O(g)$ if $\limsup_T f(T)/g(T) < \infty$ and $f = o(g)$ if $\lim_T f(T)/g(T) = 0$. We sometimes write $f = O_T(g)$ and $f = o_T(g)$ in order to clarify the parameter T going to infinity. The notation $f(T) \sim g(T)$ means that $\lim_{T\to\infty} f(T)/g(T) = 1$.

Theorem 4.4. [12] *Suppose that Γ is a geometrically finite subgroup of $\mathrm{SO}(n,1)^\circ$ with $n \geq 2$. Suppose that $\delta > (n-1)/2$ if $n = 2, 3$ and that $\delta > n-2$ if $n \geq 4$. Then there exists $\varepsilon_0 > 0$ such that for any functions $\Psi_1, \Psi_2 \in C_c^\infty(\Gamma\backslash G)$, as $t \to +\infty$,*

$$e^{(n-1-\delta)t} \int_{\Gamma\backslash G} \Psi_1(ga_t)\Psi_2(g)dg = \frac{m^{\mathrm{BR}}(\Psi_1)\cdot m_*^{\mathrm{BR}}(\Psi_2)}{|m^{\mathrm{BMS}}|} + O(e^{-\varepsilon_0 t})$$

where the implied constant depends only on the Sobolev norms of Ψ_1 and Ψ_2.

Let $\Omega \subset M$, $\Xi_1 \subset N^+$ and $\Xi_2 \subset N^-$ be bounded Borel subsets. For $T > 0$, set

$$\mathcal{S}_T(\Xi_1, \Xi_2, \Omega) = \Xi_1 A_T^+ \Omega \Xi_2. \tag{4.5}$$

By $\mathrm{Vol}(\Omega)$, we mean the volume of Ω computed with respect to the probability Haar measure on M.

Theorem 4.6. *Fix $g_0 \in G$. If $\nu_o(\partial(\Xi_1 v_o^+)) = 0 = \nu_o(\partial(\Xi_2^{-1} v_o^-))$ and $\mathrm{Vol}(\partial(\Omega)) = 0$, then as $T \to \infty$,*

$$\#\Gamma \cap g_0 \mathcal{S}_T(\Xi_1, \Xi_2, \Omega) g_0^{-1} \sim \frac{\nu_{g_0(o)}(g_0 \Xi_1 v_o^+)\nu_{g_0(o)}(g_0 \Xi_2^{-1} v_o^-)\,\mathrm{Vol}(\Omega)}{\delta |m^{\mathrm{BMS}}|} e^{\delta T}.$$

Under the assumption of Theorem 4.4, we will prove an effective version of Theorem 4.6. As usual, in order to state a result which is effective, we need to assume certain regularity condition on the boundaries of the sets Ξ_1, Ξ_2, Ω involved.

Definition 4.7. A Borel subset $\Theta \subset \partial(\tilde{X})$ is called *admissible* with respect to ν_o if there exists $r > 0$ such that for all small $\rho > 0$,

$$\nu_o\{\xi \in \partial(\tilde{X}) : d_o(\xi, \partial(\Theta)) \le \rho\} \ll \rho^r$$

Remark 4.8. In the group $G = \mathrm{SO}(2,1)^\circ$, the boundary $\partial(\tilde{X})$ is a circle, and any interval of $\partial(\tilde{X})$ is admissible. For $G = \mathrm{SO}(n,1)^\circ$ with $n \ge 3$, if $\delta > \max\{n-2, (n-2+\kappa)/2\}$ where κ is the maximum rank of parabolic fixed points of Γ, then any Borel subset ω of $\partial(\tilde{X})$ such that $\nu_o(\omega) > 0$ and $\partial(\omega)$ is a finite union of smooth sub manifolds is admissible; this is proved in [12], using Sullivan's shadow lemma.

Theorem 4.9. *Let G and Γ be as in Theorem 4.4. Suppose that $\Xi_1 v_o^+$ and $\Xi_2^{-1} v_o^-$ are admissible, and that $\partial(\Omega)$ is a finite union of smooth submanifolds. Then for any $g_0 \in G$, there exists $\varepsilon_0 > 0$ such that as $T \to \infty$,*

$$\#\Gamma \cap g_0 \mathcal{S}_T(\Xi_1, \Xi_2, \Omega) g_0^{-1} = \frac{\nu_{g_0(o)}(g_0 \Xi_1 v_o^+) \nu_{g_0(o)}(g_0 \Xi_2^{-1} v_o^-) \operatorname{Vol}(\Omega)}{\delta |m^{\mathrm{BMS}}|} e^{\delta T} + O(e^{(\delta - \varepsilon_0)T}).$$

The rest of this section is devoted to the proof of Theorems 4.6 and 4.9. In the case when $G = \mathrm{SO}(n,1)^\circ$, an analogous theorem for bisectors in KA^+K was proved in [12] (see also [3] for $n = 2$, [23] for $n = 3$ when δ is big and [9] when Γ is a lattice). In view of Theorem 4.3 for a general rank one homogeneous space admitting a finite BMS measure, the proof of Theorem 4.6 is very similar to the one given in [12] in principle.

For simplicity, we normalize $|m^{\mathrm{BMS}}| = 1$ by replacing ν_o by a suitable scalar multiple. For a given compact subset $B \subset G$, consider the following function on $\Gamma\backslash G \times \Gamma\backslash G$:

$$F_B(g,h) := \sum_{\gamma \in \Gamma} 1_B(g^{-1}\gamma h).$$

Note that for $\Psi_1, \Psi_2 \in C_c(\Gamma\backslash G)$

$$\langle F_B, \Psi_1 \otimes \Psi_2 \rangle_{\Gamma\backslash G \times \Gamma\backslash G} := \int_{\Gamma\backslash G \times \Gamma\backslash G} F_B(g_1, g_2) \Psi_1(g_1) \Psi_2(g_2) dg_1 dg_2.$$

By a standard folding and unfolding argument, we have

$$\langle F_B, \Psi_1 \otimes \Psi_2 \rangle = \int_{g \in B} \langle \Psi_1, g.\Psi_2 \rangle_{L^2(\Gamma\backslash G)} \, dg.$$

Let $\psi^\varepsilon \in C^\infty(G)$ be an ε-approximation function of e, i.e., ψ^ε is a non-negative smooth function supported on $G_\varepsilon(e)$ and $\int \psi^\varepsilon dg = 1$, and let $\Psi^\varepsilon \in C^\infty(\Gamma\backslash G)$ be its Γ-average: $\Psi^\varepsilon(\Gamma g) = \sum_{\gamma \in \Gamma} \psi^\varepsilon(\gamma g)$.

We deduce that

$$\begin{aligned}
&\langle F_B, \Psi^\varepsilon \otimes \Psi^\varepsilon \rangle_{\Gamma\backslash G \times \Gamma\backslash G} \\
&= \int_{x\in B} \int_{\Gamma\backslash G} \Psi^\varepsilon(g)\Psi^\varepsilon(gx)dgdx \\
&\text{writing } x = n_1 a_t m n_2 \in N^+AMN^- \text{ and using } dx = e^{Dt}dn_1dtdmdn_2 \\
&= \int_{n_1a_tmn_2\in B} \left(\int_{\Gamma\backslash G} \Psi^\varepsilon(g)\Psi^\varepsilon(gn_1a_tmn_2)dg \right) e^{Dt}dtdn_1dmdn_2 \\
&= \int_{n_1a_tmn_2\in B} \left(\int_{\Gamma\backslash G} \Psi^\varepsilon(gn_1^{-1})\Psi^\varepsilon(ga_tmn_2)dg \right) e^{Dt}dtdn_1dmdn_2 \\
&\text{by applying Theorem 4.3} \\
&= \int_{n_1a_tmn_2\in B} e^{\delta t}(1+o(1))m_*^{\mathrm{BR}}(n_1^{-1}\Psi^\varepsilon)m^{\mathrm{BR}}((mn_2)\Psi^\varepsilon)dtdn_1dmdn_2 \\
&= \int_{n_1a_tmn_2\in B} e^{\delta t}(1+o(1))\tilde{m}_*^{\mathrm{BR}}(n_1^{-1}\psi^\varepsilon)\tilde{m}^{\mathrm{BR}}(mn_2\psi^\varepsilon)dtdn_1dmdn_2. \quad (4.10)
\end{aligned}$$

If we define a function f_B on $N^+ \times MN^-$ by

$$f_B(n_1, mn_2) = \int_{a_t \in n_1^{-1}Bn_2^{-1}m^{-1}\cap A^+} e^{\delta t}dt,$$

and a function on $G \times G$ by

$$\begin{aligned}
&((\psi^\varepsilon \otimes \psi^\varepsilon) * f_B)(g,h) \\
&\quad = \int_{n_1mn_2\in N^+MN^-} \psi^\varepsilon(gn_1^{-1})\psi^\varepsilon(hmn_2)f_B(n_1,mn_2)dmdn_1dn_2 \\
&\quad\quad = \int_{n_1mn_2\in N^+MN^-} \psi^\varepsilon(gn_1)\psi^\varepsilon(hmn_2)f_B(n_1^{-1},mn_2)dmdn_1dn_2,
\end{aligned}$$

then we may write

$$\begin{aligned}
&\langle F_B, \Psi^\varepsilon \otimes \Psi^\varepsilon \rangle_{\Gamma\backslash G\times\Gamma\backslash G} \quad (4.11)\\
&= (\tilde{m}_*^{\mathrm{BR}} \otimes \tilde{m}^{\mathrm{BR}})((\psi^\varepsilon\otimes\psi^\varepsilon) * f_B) + o(\max_{n_1a_tmn_2\in B} e^{\delta t}).
\end{aligned}$$

Observe that

$$(\tilde{m}_*^{\mathrm{BR}} \otimes \tilde{m}^{\mathrm{BR}})((\psi^\varepsilon\otimes\psi^\varepsilon) * f_B) =$$

$$\int_{N^+MN^-} f_B(n_1^{-1}, mn_2) \left(\int_{G\times G} \psi^\varepsilon(g_1n_1)\psi^\varepsilon(h_1mn_2)d\tilde{m}_*^{\mathrm{BR}}(g_1)d\tilde{m}^{\mathrm{BR}}(h_1) \right) dn_1dmdn_2. \quad (4.12)$$

By (4.1) and (4.2), we have

$$d\tilde{m}_*^{\mathrm{BR}}(g_1)d\tilde{m}^{\mathrm{BR}}(h_1) = e^{\delta(r-r_0)}dndrdm_1d\nu_o(kv_o^+)dn_0dr_0dm_0d\nu_o(k_0v_o^-). \quad (4.13)$$

for $g_1 = km_1a_rn \in (K/M)MAN^-$ and $h_1 = k_0m_0a_{r_0}n_0 \in (K/M)MAN^+$.

For $x \in G$, let $\mathfrak{n}_1(x)$ be the N^+ component of x in MAN^-N^+ decomposition and $\tilde{\mathfrak{n}_2}(x)$ be the MN^- component of x in $AN^+(MN^-)$ decomposition. The A-components of x in MAN^-N^+ and AN^+MN^- decompositions are respectively denoted by $I_1(x)$ and $I_2(x)$.

Continuing (4.12), first change the inner integral using (4.13) and then perform the change of variables by putting $g = m_1 a_r n n_1 \in MAN^-N^+$ and $h = a_{r_0} n_0 m n_2 \in AN^+MN^-$. Since $dg = dm_1 dr dn dn_1$ and $dh = dr_0 dn_0 dm dn_2$, we obtain

$$(\tilde{m}_*^{\mathrm{BR}} \otimes \tilde{m}^{\mathrm{BR}})((\psi^\varepsilon \otimes \psi^\varepsilon) * f_B) = \tag{4.14}$$

$$\int_{k \in K/M, k_0 \in K/M, m_0 \in M} \int_{G \times G} \psi^\varepsilon(kg)\psi^\varepsilon(k_0 m_0 h) f_B(\mathfrak{n}_1(g)^{-1}, \tilde{\mathfrak{n}}_2(h)) e^{\delta(I_1(g) - I_2(h))}$$
$$dg dh d\nu_o(k v_o^+) d\nu_o(k_0 v_o^-) dm_0$$
$$= \int_{k \in K/M, k_0 \in K} \int_{G \times G} \psi^\varepsilon(g)\psi^\varepsilon(h) f_B(\mathfrak{n}_1(k^{-1}g)^{-1}, \tilde{\mathfrak{n}}_2(k_0^{-1}h))$$
$$e^{\delta(I_1(k^{-1}g) - I_2(k_0^{-1}h))} dg dh d\nu_o(k v_o^+) d\nu_o(k_0).$$

where $d\nu_o(k_0) := d\nu_o(k_0' v_o^-) dm$ for $k_0 = k_0' \times m \in K/M \times M$.

In order to prove Theorem 4.6, we now put

$$\mathcal{S}_T := \mathcal{S}_T(\Xi_1, \Xi_2, \Omega) \text{ and } F_T := F_{\mathcal{S}_T}.$$

Observe that

$$F_T(e, e) = \#(\Gamma \cap \mathcal{S}_T(\Xi_1, \Xi_2, \Omega)).$$

Let

$$\mathcal{S}_{T,\varepsilon}^+ = \cup_{g_1, g_2 \in G_\varepsilon(e)} g_1 \mathcal{S}_T g_2 \quad \text{and} \quad \mathcal{S}_{T,\varepsilon}^- = \cap_{g_1, g_2 \in G_\varepsilon(e)} g_1 \mathcal{S}_T g_2.$$

We then have

$$\langle F_{\mathcal{S}_{T,\varepsilon}^-}, \Psi^\varepsilon \otimes \Psi^\varepsilon \rangle \le F_T(e, e) \le \langle F_{\mathcal{S}_{T,\varepsilon}^+}, \Psi^\varepsilon \otimes \Psi^\varepsilon \rangle. \tag{4.15}$$

Together with the strong wave front property for the $AN^\pm K$ decompositions [10], (4.14) with $B = \mathcal{S}_{T,\varepsilon}^\pm$ now implies that

$$(\tilde{m}_*^{\mathrm{BR}} \otimes \tilde{m}^{\mathrm{BR}})((\psi^\varepsilon \otimes \psi^\varepsilon) * f_{\mathcal{S}_{T,\varepsilon}^\pm}) \tag{4.16}$$
$$= (1 + O(\varepsilon')) \int_{K/M \times K} f_{\mathcal{S}_T}(\mathfrak{n}_1(k^{-1})^{-1}, \tilde{\mathfrak{n}}_2(m^{-1}k_0^{-1})) d\nu_o(k v_o^+) d\nu_o(k_0)$$
$$= (1 + O(\varepsilon')) \frac{e^{\delta T}}{\delta} \nu_o(\Xi_1 v_o^+) \nu_o(\Xi_2^{-1} v_o^-) \operatorname{Vol}(\Omega)$$

where $\varepsilon' > 0$ goes to 0 as $\varepsilon \to 0$. Hence, (4.15), (4.10) and (4.11) yield that

$$F_T(e, e) = (1 + O(\varepsilon')) \frac{e^{\delta T}}{\delta} \nu_o(\Xi_1 v_o^+) \nu_o(\Xi_2^{-1} v_o^-) \operatorname{Vol}(\Omega) + o(e^{\delta T}).$$

Since $\varepsilon > 0$ is arbitrary, we have

$$F_T(e,e) \sim \frac{e^{\delta T}}{\delta}\nu_o(\Xi_1 v_o^+)\nu_o(\Xi_2^{-1}v_o^-)\operatorname{Vol}(\Omega).$$

In order to prove Theorem 4.9 for $g_0 = e$, we note that $o(e^{\delta T})$ in (4.10) can be upgraded into $O(e^{(\delta-\varepsilon_0)T})$ in view of Theorem 4.4, and that $O(\varepsilon')$ in (4.16) can be taken as $O(\varepsilon^q)$ for some fixed $q > 0$ (we refer to [12] for details).

Therefore we get

$$F_T(e,e) = (1+O(\varepsilon^q))\frac{e^{\delta T}}{\delta}\nu_o(\Xi_1 v_o^+)\nu_o(\Xi_2^{-1}v_o^-)\operatorname{Vol}(\Omega) + O(e^{(\delta-\varepsilon_0)T}).$$

By taking ε so that $\varepsilon^q = e^{-\varepsilon_0 t}$, we then obtain

$$F_T(e,e) = \frac{e^{\delta T}}{\delta}\nu_o(\Xi_1 v_o^+)\nu_o(\Xi_2^{-1}v_o^-)\operatorname{Vol}(\Omega) + O(e^{(\delta-\varepsilon_1)T})$$

for some positive $\varepsilon_1 > 0$. This proves Theorems 4.6 and 4.9 for $g_0 = e$.

For a general $g_0 \in G$, we note that if we set $\Gamma_0 := g_0^{-1}\Gamma g_0$, then

$$\#\Gamma \cap g_0\mathcal{S}_T(\Xi_1,\Xi_2,\Omega)g_0^{-1} = \#\Gamma_0 \cap \mathcal{S}_T(\Xi_1,\Xi_2,\Omega).$$

Moreover, if we set $\nu_{\Gamma_0,x} := g_0^*\nu_{g_0(x)}$, then $\{\nu_{\Gamma_0,x} : x \in \tilde{X}\}$ is a Γ_0-invariant conformal density of dimension $\delta = \delta_{\Gamma_0}$, and the corresponding BMS-measure $m^{\mathrm{BMS}}_{\Gamma_0}$ with respect to $\{\nu_{\Gamma_0,x}\}$ has the same total mass as m^{BMS}. Therefore

$$\frac{\nu_{\Gamma_0,o}(\Xi_1 v_o^+)\nu_{\Gamma_0,o}(\Xi_2^{-1}v_o^-)}{\delta_{\Gamma_0}|m^{\mathrm{BMS}}_{\Gamma_0}|} = \frac{\nu_{g_0(o)}(g_0\Xi_1 v_o^+)\nu_{g_0(o)}(g_0\Xi_2^{-1}v_o^-)}{\delta|m^{\mathrm{BMS}}|}.$$

Hence the general case follows from $g_0 = e$.

4.2. **On the counting for $\Gamma \cap \mathfrak{B}(\varepsilon)A_T^+\Omega\mathfrak{B}(\varepsilon)^{-1}$.** Recall the definition of our flow box at $g_0 \in G$ with $\varepsilon > 0$ smaller than the injectivity radius of g_0 in $\Gamma\backslash G$:

$$\mathfrak{B}(g_0,\varepsilon) = g_0(N_\varepsilon^+N^- \cap N_\varepsilon^-N^+AM)M_\varepsilon A_\varepsilon. \tag{4.17}$$

Denote $\tilde{\pi} : G \to \Gamma\backslash G/M$ the canonical projection map. For simplicity, we set

$$\tilde{\mathfrak{B}}(g_0,\varepsilon) = \tilde{\pi}(\mathfrak{B}(g_0,\varepsilon)). \tag{4.18}$$

For a Borel function f on $\Gamma\backslash G/M$ and a Borel function ξ on M, we set

$$m^{\mathrm{BMS}}(f\otimes\xi) := \int_{\mathrm{T}^1(X)} f\,dm^{\mathrm{BMS}} \cdot \int_M \xi\,dm;$$

For Borel subsets $B \subset \Gamma\backslash G/M$ and $\Omega \subset M$, we set $m^{\mathrm{BMS}}(B\otimes\Omega) = m^{\mathrm{BMS}}(1_B \otimes 1_\Omega)$. We observe that:

Lemma 4.19. *For all small $\varepsilon > 0$,*

$$m^{\mathrm{BMS}}(\tilde{\mathfrak{B}}(g_0,\varepsilon)\otimes\Omega) = (1+O(\varepsilon))2\varepsilon\cdot\nu_{g_0(o)}((g_0N_\varepsilon^+)v_o^+)\nu_{g_0(o)}((g_0N_\varepsilon^-)v_o^-)\operatorname{Vol}(\Omega)$$

where the implied constant is independent of $\varepsilon > 0$.

Proof. Clearly we have $m^{\mathrm{BMS}}(\tilde{\mathfrak{B}}(g_0,\varepsilon)\otimes\Omega) = \tilde{m}^{\mathrm{BMS}}(\mathfrak{B}(g_0,\varepsilon)\otimes\Omega)$. Recall that the BMS measure on $\mathrm{T}^1(\tilde{X})$ is given as

$$d\tilde{m}^{\mathrm{BMS}}(u) = \frac{d\nu_{g_0(o)}(u^+)d\nu_{g_0(o)}(u^-)ds}{d_{g_0(o)}(u^+,u^-)^{2\delta}}.$$

Note that

$$\mathfrak{B}(g_0,\varepsilon)v_o^+ = g_0 N_\varepsilon^+ v_o^+$$

(which is equal to the image of $\mathfrak{B}(g_0,\varepsilon)$ in $G/(MAN^-)$) and

$$\mathfrak{B}(g_0,\varepsilon)v_o^- = g_0 N_\varepsilon^- v_o^-$$

(which is equal to the image of $\mathfrak{B}(g_0,\varepsilon)$ in $G/(MAN^+)$) . Hence for all $g\in\mathfrak{B}(g_0,\varepsilon)$, we have $d_{g_0(o)}(g^+,g^-) = (1+O(\varepsilon))$ where the implied constant is independent of $g_0\in G$ and $\varepsilon>0$. Moreover, for all $g\in\mathfrak{B}(g_0,\varepsilon)$, $\{t\in\mathbb{R} : ga_t\in\mathfrak{B}(g_0,\varepsilon)\}$ has length precisely 2ε (see Lemma 2.4). Therefore the claim follows, since the BMS measure on G is the M-invariant extension of the BMS measure of G/M. $\square$

For $T>1$ and $g_0\in G$, we define

$$\mathcal{V}_T(g_0,\varepsilon,\Omega) := \mathfrak{B}(g_0,\varepsilon)A_T^+\Omega\mathfrak{B}(g_0,\varepsilon)^{-1}. \tag{4.20}$$

We set

$$\mathcal{V}_T(\varepsilon,\Omega) := \mathcal{V}_T(e,\varepsilon,\Omega)$$

and note that

$$\mathcal{V}_T(g_0,\varepsilon,\Omega) := g_0\mathcal{V}_T(\varepsilon,\Omega)g_0^{-1}.$$

Lemma 4.21. *For all large $T\gg 1$ and small $0<\varepsilon<1$, we have*

$$\mathcal{S}_T(N_\varepsilon^+,(N_\varepsilon^-)^{-1},\Omega)\subset\mathcal{V}_T(\varepsilon,\Omega)\subset\mathcal{S}_{T+\varepsilon}(N^+_{\varepsilon+e^{-T}},(N^-_{\varepsilon-e^{-T}})^{-1},\Omega_\varepsilon^+)$$

where $\Omega_\varepsilon^+ = \cup_{m_i\in M_\varepsilon} m_1\Omega m_2$.

Proof. Given $g\in\mathfrak{B}(\varepsilon)\cup\mathfrak{B}(\varepsilon)^{-1}$, we decompose

$$g = g_+g_0g_-\in N^+(AM)N^-.$$

It easily follows from the definition of $\mathfrak{B}(\varepsilon)$ that

$$N_\varepsilon^+ = \{g_+ : g\in\mathfrak{B}(\varepsilon)\}\quad\text{ and }\quad (N_\varepsilon^-)^{-1} = \{g_- : g\in\mathfrak{B}(\varepsilon)^{-1}\}.$$

Hence

$$\mathcal{S}_T(N_\varepsilon^+,(N_\varepsilon^-)^{-1},\Omega)\subset\mathcal{V}_T(\varepsilon,\Omega). \tag{4.22}$$

On the other hand, if $g_1\in\mathfrak{B}(\varepsilon)$, $g_2\in\mathfrak{B}(\varepsilon)^{-1}$, $a\in A_T^+$, and $m\in M$, then

$$g_1amg_2\in(g_1)_+N^+_{e^{-T}}amA_\varepsilon M_\varepsilon(N^-_{e^{-T}})^{-1}(g_2)_-.$$

Therefore

$$\mathcal{V}_T(\varepsilon,\Omega)\subset\mathcal{S}_{T+\varepsilon}(N^+_{\varepsilon+e^{-T}},(N^-_{\varepsilon-e^{-T}})^{-1},\Omega_\varepsilon^+). \tag{4.23}$$

This proves the claim. $\square$

Theorem 4.24. *Let $\varepsilon > 0$ be smaller than the injectivity radius of g_0. We have*

$$\#\Gamma \cap \mathcal{V}_T(g_0, \varepsilon, \Omega) = (1 + O(\varepsilon)) \frac{e^{\delta T}}{\delta \cdot 2\varepsilon \cdot |m^{\mathrm{BMS}}|} \cdot (m^{\mathrm{BMS}}(\mathfrak{B}(g_0, \varepsilon) \otimes \Omega) + o(1))$$

where the implied constants are independent of ε.

Moreover if G and Γ are as in Theorem 1.2, $o(1)$ can be replaced by $O(e^{-\varepsilon_1 T})$ for some positive $\varepsilon_1 > 0$.

Proof. By Lemma 4.21, we have

$$g_0 \mathcal{S}_T(N_\varepsilon^+, (N_\varepsilon^-)^{-1}, \Omega) g_0^{-1} \subset \mathcal{V}_T(g_0, \varepsilon, \Omega) \subset g_0 \mathcal{S}_T(N^+_{\varepsilon+e^{-T}}, (N^-_{\varepsilon-e^{-T}})^{-1}, \Omega_\varepsilon^+) g_0^{-1}.$$

By Theorem 4.6 and Lemma 4.19, we have

$$\begin{aligned}
&\#\Gamma \cap \mathcal{V}_T(g_0, \varepsilon, \Omega) \\
&\quad = (1 + o(1)) \cdot \nu_{g_0(o)}((g_0 N_\varepsilon^+) v_o^+) \nu_{g_0(o)}((g_0 N_\varepsilon^-) v_o^-) \operatorname{Vol}(\Omega) \delta^{-1} e^{\delta T} \\
&\qquad\quad = (1 + O(\varepsilon))(2\varepsilon)^{-1} \delta^{-1} e^{\delta T} \cdot (\tilde{m}^{\mathrm{BMS}}(\mathfrak{B}(g_0, \varepsilon) \otimes \Omega) + o(1)),
\end{aligned}$$

implying the first claim. The second claim follows from Theorem 4.9, and Remark 4.8. $\square$

5. Asymptotic distribution of closed geodesics with holonomies

We keep the notations G, Γ, X, K, o, v_o etc. from section 3. In particular, Γ is Zariski dense and $|m^{\mathrm{BMS}}| < \infty$, $X = \Gamma \backslash G/K$, and $\mathrm{T}^1(X) = \Gamma \backslash G/M$. In this section, we will describe the distribution of all closed geodesics of length at most T coupled together with their holonomy classes, using the results proved in section 4. The main ingredient is the comparison lemma 5.12, which we obtain using the effective closing lemma 3.7.

By a (primitive) closed geodesic C in $\mathrm{T}^1(X)$, we mean a compact set of the form

$$\Gamma \backslash \Gamma g AM/M = \Gamma \backslash \Gamma g A(v_o)$$

for some $g \in G$. The length of a closed geodesic $C = \Gamma \backslash \Gamma g AM/M$ is same as the co-volume of $AM \cap g^{-1} \Gamma g$ in AM. If we denote by γ_C a generator of $\Gamma \cap gAMg^{-1}$ and denote by $[\gamma_C]$ its conjugacy class in Γ, then the map

$$C \mapsto [\gamma_C]$$

is a bijection between between the set of all (primitive) closed geodesics and the set of all primitive hyperbolic conjugacy classes of Γ.

For each closed geodesic C, we denote by $\mathcal{L}_C$ the length measure on C and by h_C the unique M-conjugacy class associated to the holonomy class of C. For a primitive hyperbolic element $\gamma \in \Gamma$, we denote by $\ell(\gamma)$ its translation length, or equivalently the length of the closed geodesic corresponding to $[\gamma]$.

Let M^{c} denote the space of conjugacy classes of M. It is known that M^{c} can be identified with $\mathrm{Lie}(S)/W$ where S is a maximal torus of M and W is the Weyl group relative to S. For $T > 0$, define

$$\mathcal{G}_\Gamma(T) := \{C : C \text{ is a closed geodesic in } \mathrm{T}^1(X),\ \ \ell(C) \le T\}.$$

For each $T > 0$, we define the measure μ_T on the product space $(\Gamma\backslash G/M)\times M^{\mathrm{c}}$: for $f \in C(\Gamma\backslash G/M)$ and any class function $\xi \in C(M)$,

$$\mu_T(f\otimes\xi) = \sum_{C\in\mathcal{G}_\Gamma(T)} \mathcal{L}_C(f)\xi(h_C).$$

We also define a measure η_T by

$$\eta_T(f\otimes\xi) = \sum_{C\in\mathcal{G}_\Gamma(T)} \mathcal{D}_C(f)\xi(h_C),$$

where $\mathcal{D}_C(f) = \ell(C)^{-1}\mathcal{L}_C(f)$. If B is a subset of $\Gamma\backslash G/M$ and Ω is a subset of M, then we put $\mu_T(B\otimes\Omega) := \mu_T(1_B\otimes 1_\Omega)$ and $\eta_T(B\otimes\Omega) := \eta_T(1_B\otimes 1_\Omega)$.

The main goal of this section is to prove the following:

Theorem 5.1. *Let Γ be geometrically finite and Zariski dense. For any bounded $f \in C(\Gamma\backslash G/M)$ and $\xi \in \mathrm{Cl}(M)$, we have, as $T \to \infty$,*

$$\mu_T(f\otimes\xi) \sim \frac{e^{\delta T}}{\delta|m^{\mathrm{BMS}}|}\cdot m^{\mathrm{BMS}}(f\otimes\xi); \tag{5.2}$$

and

$$\eta_T(f\otimes\xi) \sim \frac{e^{\delta T}}{\delta\cdot|m^{\mathrm{BMS}}|\cdot T}\cdot m^{\mathrm{BMS}}(f\otimes\xi). \tag{5.3}$$

Moreover if G and Γ are as Theorem 1.2, then both (5.2) *and* (5.3) *hold with an exponential error term $O(e^{-\varepsilon_1}t)$ for some $\varepsilon_1 > 0$ with the implied constants depending only on the Sobolev norms of f and ξ.*

Theorems 1.1 and 1.2 in the introduction follow immediately from Theorem 5.1, whose proof occupies the rest of this section.

Fix a Borel subset Ω of M^{c} and $g_0 \in G$. Recall the flow box $\mathfrak{B}(g_0,\varepsilon) = g_0(N_\varepsilon^+N^-\cap N_\varepsilon^-N^+AM)M_\varepsilon A_\varepsilon$ and the notation $\tilde{\mathfrak{B}}(g_0,\varepsilon) = \tilde\pi(\mathfrak{B}(g_0,\varepsilon))$ from (4.17) and (4.18). We will first investigate the measure μ_T restricted to the set $\tilde{\mathfrak{B}}(g_0,\varepsilon)\otimes\Omega$. The main idea is to relate the measure $\mu_T(\tilde{\mathfrak{B}}(g_0,\varepsilon)\otimes\Omega)$ with the cardinality $\#\Gamma\cap\mathcal{V}_T(g_0,\varepsilon,\Omega)$.

We fix $g_0 \in G$ and $\varepsilon > 0$ (smaller than the injectivity radius of g_0) from now on until Theorem 5.14. For a closed geodesic $C = \Gamma\backslash\Gamma gAv_o \subset \Gamma\backslash G/M$, we choose a complete geodesic $\tilde C \subset G/M$, which is a lift of C. The stabilizer $\Gamma_{\tilde C} = \{\gamma\in\Gamma : \gamma(\tilde C) = \tilde C\}$ is $gAMg^{-1}\cap\Gamma$ which is generated by a primitive hyperbolic element of Γ, and C can be identified with $\Gamma_{\tilde C}\backslash\tilde C$. Set

$$I(C) = \{[\sigma]\in\Gamma/\Gamma_{\tilde C} : \sigma\tilde C\cap\mathfrak{B}(g_0,\varepsilon)v_o \ne \emptyset\}, \tag{5.4}$$

that is, $I(C) = \{\sigma\tilde{C} : \sigma\tilde{C} \cap \mathfrak{B}(g_0, \varepsilon)v_o \neq \emptyset\}$. Clearly $\#I(C)$ does not depend on the choice of $\tilde{C}$.

Lemma 5.5. (1) *For any closed geodesic $C \subset \mathrm{T}^1(X)$, we have*

$$\mathcal{L}_C(\tilde{\mathfrak{B}}(g_0, \varepsilon)) = 2\varepsilon \cdot \#I(C);$$

(2) *For any $T > 0$, we have*

$$\mu_T(\tilde{\mathfrak{B}}(g_0, \varepsilon) \otimes \Omega) = 2\varepsilon \cdot \sum_{C \in \mathcal{G}_\Gamma(T)} \#I(C) \cdot 1_\Omega(h_C) \tag{5.6}$$

where h_C is the holonomy class about C.

Proof. (2) immediately follows from (1). To see (1), let $C = \Gamma\backslash\Gamma g A v_o$. We may assume $\tilde{C} = gAv_o$. We have

$$\begin{aligned}\mathcal{L}_C(\tilde{\mathfrak{B}}(g_0, \varepsilon)) &= \int_{[ga_t v_o] \in \Gamma_{\tilde{C}}\backslash\tilde{C}} \sum_{\sigma \in \Gamma} 1_{\mathfrak{B}(g_0,\varepsilon)}(\sigma g a_t v_o) dt \\ &= \sum_{[\sigma] \in \Gamma/\Gamma_{\tilde{C}}} \int_{g a_t v_o \in \tilde{C}} 1_{\mathfrak{B}(g_0,\varepsilon)}(\sigma g a_t v_o) dt.\end{aligned}$$

By Lemma 2.4, we have

$$\int_{\tilde{C}} 1_{\mathfrak{B}(g_0,\varepsilon)}(\sigma g a_t v_o) dt = \begin{cases} 2\varepsilon, & \text{if } \sigma\tilde{C} \cap \mathfrak{B}(g_0, \varepsilon)v_o \neq 0 \\ 0, & \text{otherwise.} \end{cases} \tag{5.7}$$

Therefore the claim follows. $\square$

Set

$$\mathcal{W}(g_0, \varepsilon, \Omega) := \{gamg^{-1} : g \in \mathfrak{B}(g_0, \varepsilon), am \in A\Omega\}.$$

By definition, the set $\mathcal{W}(g_0, \varepsilon, \Omega)$ consists of hyperbolic elements. For $T > 1$, we set

$$\mathcal{W}_T(g_0, \varepsilon, \Omega) := \{gamg^{-1} : g \in \mathfrak{B}(g_0, \varepsilon), am \in A_T^+\Omega\}.$$

We denote by Γ_h the set of hyperbolic elements and by Γ_{ph} the set of primitive hyperbolic elements of Γ.

Proposition 5.8. *For all large $T \gg 1$, we have*

$$\mu_T(\tilde{\mathfrak{B}}(g_0, \varepsilon) \otimes 1_\Omega) = 2\varepsilon \cdot \#\Gamma_{ph} \cap \mathcal{W}_T(g_0, \varepsilon, \Omega)$$

Proof. We use Lemma 5.5 (2):

$$\mu_T(\tilde{\mathfrak{B}}(g_0, \varepsilon) \otimes 1_\Omega) = 2\varepsilon \cdot \sum_{C \in \mathcal{G}_\Gamma(T)} \#I(C) \cdot 1_\Omega(h_C).$$

with $I(C) = \{\sigma(\tilde{C}) : \sigma\tilde{C} \cap \mathfrak{B}(g_0, \varepsilon)v_o \neq \emptyset\}$.

Upper bound: Let $C \in \mathcal{G}_\Gamma(T)$ be with $I(C)$ non-empty and $h_C \in \Omega$. Without loss of generality, we may assume $\tilde{C} \cap \mathfrak{B}(g_0, \varepsilon)v_o \neq \emptyset$. Choose a primitive hyperbolic element $\gamma := \gamma_C \in \Gamma_{\tilde{C}}$. We claim that for any $[\sigma] \in I(C)$,

$$\sigma_\gamma := \sigma\gamma\sigma^{-1} \in \mathcal{W}_T(g_0, \varepsilon, \Omega); \tag{5.9}$$

note that σ_γ is well-defined independent of the choice of a representative σ since $\Gamma_{\tilde{C}}$ is commutative. Since $\tilde{C} \cap \mathfrak{B}(g_0, \varepsilon) v_o \neq \emptyset$, there exists $g_1 \in \mathfrak{B}(g_0, \varepsilon)$ such that $g_1 v_o \in \tilde{C}$, and $\gamma = g_1 a_\gamma m_\gamma g_1^{-1}$ where $d(a_\gamma, e) = \ell(C) \leq T$ and $[m_\gamma] \in \Omega$. If $[\sigma] \in I(C)$, then there exist $g_2 \in \mathfrak{B}(g_0, \varepsilon)$ and $a_s m \in AM$ such that

$$\sigma g_1 a_s m = g_2.$$

Therefore, we have

$$g_2 a_\gamma m^{-1} m_\gamma m = \sigma \gamma \sigma^{-1} g_2$$

and

$$\sigma_\gamma = g_2 a_\gamma m^{-1} m_\gamma m g_2^{-1} \in \mathcal{W}_T(g_0, \varepsilon, \Omega).$$

proving (5.9).

To see that the map $[\sigma] \mapsto \sigma_\gamma$ is injective on $I(C)$, it suffices to recall that the centralizer of γ in Γ is $\Gamma_{\tilde{C}}$. Hence this proves the upper bound.

Lower bound: We write

$$\# \Gamma_{ph} \cap \mathcal{W}_T(g_0, \varepsilon, \Omega) = \sum \# [\gamma] \cap \mathcal{W}_T(g_0, \varepsilon, \Omega)$$

where the sum ranges over the conjugacy classes

$$[\gamma] = \{\gamma_0 \in \Gamma_{ph} : \gamma_0 \text{ is conjugate to } \gamma \text{ by an element of } \Gamma\}$$

of primitive hyperbolic elements of Γ. Fix a primitive hyperbolic clement $\gamma \in \mathcal{W}_T(g_0, \varepsilon, \Omega)$. So there exists $g \in \mathfrak{B}(g_0, \varepsilon)$ such that $\gamma = g a_\gamma m_\gamma g^{-1}$ with $a_\gamma \in A_T^+$ and $[m_\gamma] \in \Omega$. Let $C = \Gamma \backslash \Gamma g a A v_o$ and $\tilde{C} = g A v_o$. Then the length of C is at most T.

For each element $\sigma' := \sigma \gamma \sigma^{-1} \in [\gamma] \cap \mathcal{W}_T(g_0, \varepsilon, \Omega)$, we have $\sigma \gamma \sigma^{-1} = g_2 a_\gamma m g_2^{-1}$ for some $[m] \in \Omega$ and $g_2 \in \mathfrak{B}(g_0, \varepsilon)$.

Since $\sigma^{-1} g_2 A v_o$ is the oriented axis for γ, $\sigma^{-1} g_2 A v_o = \tilde{C}$ by Corollary 2.7. Therefore $g_2 v_o \in \sigma(\tilde{C}) \cap \mathfrak{B}(g_0, \varepsilon) v_o$, and hence $\sigma(\tilde{C}) \in I(C)$. Since the map $\sigma' = \sigma \gamma \sigma^{-1} \mapsto \sigma(\tilde{C})$ is well-defined and injective, this proves the lower bound by (5.6). $\square$

Indeed the proof of Proposition 5.8 gives that if C is a closed geodesic and $[\gamma_C]$ is the conjugacy class of primitive hyperbolic elements which corresponds to C, then

$$\mathcal{L}_C(\tilde{\mathfrak{B}}(g_0, \varepsilon)) 1_\Omega(h_C) = 2\varepsilon \cdot \# I(C) \cdot 1_\Omega(h_C) = 2\varepsilon \cdot \# [\gamma_C] \cap \mathcal{W}(g_0, \varepsilon, \Omega). \quad (5.10)$$

Lemma 5.11. *For $T > 1$, we have*

$$\# \Gamma \cap \left(\mathcal{W}_T(g_0, \varepsilon, \Omega) - \mathcal{W}_{T/2}(g_0, \varepsilon, \Omega) \right) \leq \# \Gamma_{ph} \cap \mathcal{W}_T(g_0, \varepsilon, \Omega).$$

Proof. Note that, since $\mathcal{W}_T(g_0, \varepsilon, \Omega)$ consists of hyperbolic elements,

$$\# \Gamma_{ph} \cap \mathcal{W}_T(g_0, \varepsilon, \Omega) = \# \Gamma \cap \mathcal{W}_T(g_0, \varepsilon, \Omega) - \# (\cup_{k \geq 2} \Gamma_{ph}^k) \cap \mathcal{W}_T(g_0, \varepsilon, \Omega)$$

where $\Gamma_{ph}^k = \{\sigma^k : \sigma \in \Gamma_{ph}\}$.

Suppose $\gamma \in \mathcal{W}_T(g_0, \varepsilon, \Omega) \cap \Gamma_{ph}^k$ for $k \geq 2$, so that $\gamma = \sigma^k$ for a unique element $\sigma := \sigma_\gamma \in \Gamma_{ph}$. Since $\gamma \in \mathcal{W}_T(g_0, \varepsilon, \Omega)$, there exists $g \in \mathfrak{B}(g_0, \varepsilon)$

such that $\gamma = ga_\gamma m_\gamma g^{-1}$. Since γ is hyperbolic, so is σ, and γ and σ have the same oriented axis. Therefore $\sigma = ga_\sigma m_\sigma g^{-1}$. Since $\sigma^k = \gamma$, it follows that $a_\sigma^k = a_\gamma$ and hence $a_\sigma \in A^+_{T/k}$.

Therefore $\sigma = \sigma_\gamma \in \mathcal{W}_{T/2}(g_0, \varepsilon, \Omega)$. Since the map $\gamma \mapsto \sigma_\gamma$ is injective, it follows that

$$\#\mathcal{W}_T(g_0, \varepsilon, \Omega) \cap (\cup_{k\geq 2}\Gamma^k_{ph}) \leq \#\Gamma \cap \mathcal{W}_{T/2}(g_0, \varepsilon, \Omega).$$

Therefore the lower bound now follows. □

By the ergodicity of the geodesic flow with respect to the BMS measure on $\Gamma\backslash G$ [24], for any $g_0 \in \mathrm{supp}(m^{\mathrm{BMS}})$, a random AM-orbit in $\Gamma\backslash G$ comes back to the flow box $\mathfrak{B}(g_0, \varepsilon)$ infinitely often. The effective closing lemma implies that that there is an arbitrarily long closed geodesic nearby whose holonomy class is $O(\varepsilon)$-close to the M-component of g_0 in the N^+N^-AM decomposition. Since the projection of $\mathrm{supp}(m^{\mathrm{BMS}})$ to the M-components is all of M, this shows not only the existence of a closed geodesic but also the density of holonomy classes in the space of all conjugacy classes of M.

The comparison lemma below gives a much stronger control on the number of closed geodesics whose holonomy classes contained in a fixed subset of M in terms of lattice points, whose cardinality is controlled by the mixing.

Recall the notation

$$\mathcal{V}_T(g_0, \varepsilon, \Omega) := \mathfrak{B}(g_0, \varepsilon)A^+_T\Omega\mathfrak{B}(g_0, \varepsilon)^{-1}.$$

Let $c > 1$ be a fixed upper bound for all implied constants involved in the O symbol in Lemma 3.7 and the constant in (5.12).

Lemma 5.12 (Comparison Lemma). *For all $T \gg 1$, we have*

$$2\varepsilon \cdot \#\Gamma \cap \Big(\mathcal{V}_T(g_0, \varepsilon(1 - ce^{-T/2}), \Omega^-_{c\varepsilon}) - \mathcal{V}_{T/2}(g_0, \varepsilon, \Omega)\Big) \leq \mu_T(\tilde{\mathfrak{B}}(g_0, \varepsilon) \otimes \Omega) \leq 2\varepsilon \cdot \#\Gamma \cap \mathcal{V}_T(g_0, \varepsilon, \Omega).$$

where $\Omega^-_{c\varepsilon} = \cap_{m_i \in M_{c\varepsilon}} m_1\Omega m_2$.

Proof. The upper bound is immediate from the definition of the sets and Proposition 5.8. The effective closing lemma 3.1 implies that for all large $T \gg 1$,

$$\mathcal{V}_T(g_0, \varepsilon(1 - ce^{-T/2}), \Omega^-_{c\varepsilon}) - \mathcal{V}_{T/2}(g_0, \varepsilon, \Omega) \subset \mathcal{W}_T(g_0, \varepsilon, \Omega). \tag{5.13}$$

Together with Proposition 5.8 and Lemma 5.11, this implies the lower bound. □

Theorem 5.14. *We have*

$$\mu_T(\tilde{\mathfrak{B}}(g_0, \varepsilon)\otimes\Omega) = (1+O(\varepsilon))\frac{e^{\delta T}}{\delta \cdot |m^{\mathrm{BMS}}|}\cdot(m^{\mathrm{BMS}}(\tilde{\mathfrak{B}}(g_0, \varepsilon)\otimes\Omega)+o(1)) \tag{5.15}$$

where the implied constants are independent of g_0 and ε.

Moreover if G and Γ are as in Theorem 1.2, $o(1)$ can be replaced by $O(e^{-\varepsilon_1 T})$ for some positive $\varepsilon_1 > 0$.

Proof. This follows from the comparison lemma 5.12 and Theorem 4.24. □

We note that we do not require Γ to be geometrically finite in the following theorem.

Theorem 5.16. *Let Γ be Zariski dense with $|m^{\mathrm{BMS}}| < \infty$. For any $f \in C_c(\Gamma\backslash G/M)$ and $\xi \in \mathrm{Cl}(M)$, we have, as $T \to \infty$,*

$$\mu_T(f \otimes \xi) \sim \frac{e^{\delta T} \cdot m^{\mathrm{BMS}}(f \otimes \xi)}{\delta \cdot |m^{\mathrm{BMS}}|}. \tag{5.17}$$

Moreover if G and Γ are as in Theorem 1.2, then (5.17) *holds with an exponential error term $O(e^{-\varepsilon_1 t})$ for some $\varepsilon_1 > 0$ with the implied constants depending on the Sobolev norms of f and ξ.*

Proof. We normalize $|m^{\mathrm{BMS}}| = 1$. Using a partition of unity argument, we can assume without loss of generality that f is supported on $\tilde{\mathfrak{B}}(g_0, \varepsilon)$ for some $g_0 \in G$ and $\varepsilon > 0$. Now for arbitrarily small $0 < \rho < \varepsilon$, we can approximate f as step functions which are linear combination of characteristic functions of $\tilde{\mathfrak{B}}(h, \rho)$'s with $h \in \tilde{\mathfrak{B}}(g_0, \varepsilon)$. Now applying Proposition 5.14 to each $1_{\tilde{\mathfrak{B}}(h,\rho)} \otimes 1_\Omega$, we deduce that

$$\begin{aligned}(1 - c\rho) m^{\mathrm{BMS}}(f \otimes \xi) \le \liminf_T e^{-\delta T} \mu_T(f \otimes \xi) \le \\ \limsup_T \delta e^{-\delta T} \mu_T(f \otimes \xi) \le (1 + c\rho) m^{\mathrm{BMS}}(f \otimes \xi)\end{aligned}$$

Since $\rho > 0$ is arbitrary, this implies the claim when ξ is the characteristic function of Ω whose boundary has a measure zero. Via the identification $M^{\mathrm{C}} = \mathrm{Lie}(S)/W$ where S is a maximal torus of M and W is the Weyl group relative to S, extending the above claim from characteristic (class) functions to continuous (class) functions is similar to the above arguments. This establishes (5.17). When the effective version of Theorem 5.14 holds, we also obtain an error term in this argument. □

Contribution of the cusp and equidistribution for bounded functions In order to extend Theorem 5.16 to bounded continuous functions, which are not necessarily compactly supported, we now assume that Γ is geometrically finite and use the following theorem of Roblin [18] (Theorem 5.19).

We denote by $\mathcal{C}(\Gamma)$ the convex core of Γ. Let $\varepsilon_0 > 0$ be the Margulis constant for Γ. Then $\{x \in \mathcal{C}(\Gamma) : \text{injectivity radius at } x \ge \varepsilon_0\}$ is called the thick part and its complement is called the thin part. We will denote them $\mathcal{C}(\Gamma)_{thick}$ and $\mathcal{C}(\Gamma)_{thin}$ respectively. When Γ is a geometrically finite group, the thin part part consists of finitely many disjoint cuspidal regions (called horoballs), say, $\mathcal{H}_1, \cdots, \mathcal{H}_k$ based at parabolic fixed points $p_1, \cdots, p_k$ respectively. We denote by Γ_{p_i} the stabilizer of p_i in Γ. Also, fixing o in the thick part of $\mathcal{C}(\Gamma)$, let q_i denote the point of intersection between the geodesic ray connecting o and p_i with the boundary of the horoball $\mathcal{H}_i$.

Proposition 5.18. [5] *If Γ is geometrically finite, then for each parabolic fixed point $p_i \in \Lambda(\Gamma)$, we have*

$$\sum_{\sigma \in \Gamma_{p_i}} d(q_i, \sigma q_i) e^{-\delta \cdot d(q_i, \sigma q_i)} < \infty.$$

For any $r \geq 0$ denote by $\mathcal{H}_i(r)$ the horoball contained in $\mathcal{H}_i$ whose boundary is of distance r to $\partial \mathcal{H}_i$. Put $\mathrm{cusp}(r) = \cup_i \mathcal{H}_i(r)$.

Theorem 5.19 (Roblin, [18])**.** *There exist absolute constants $c_0, c_1 > 0$ such that for any $T \gg 1$,*

$$e^{-\delta T} \cdot \mu_T(\mathrm{cusp}(r)K) \leq c_1 \sum_{i=1}^{k} \sum_{\sigma \in \Gamma_{p_i}, d(q_i, \sigma q_i) > 2r - c_0} (d(q_i, \sigma q_i) - 2r + c_0) e^{-\delta \cdot d(q_i, \sigma q_i)}.$$

In particular, if $G = \mathrm{SO}(n,1)^\circ$, then

$$e^{-\delta T} \cdot \mu_T(\mathrm{cusp}(r)K) \ll e^{(\kappa - 2\delta) r} \tag{5.20}$$

where $\kappa = \max \mathrm{rank}(p_i)$.

These estimates and the proof for compactly supported functions imply the result for bounded functions.

Proof of Theorem 5.1: We may assume $|m^{\mathrm{BMS}}| = 1$. By Proposition 5.18,

$$\sum_{\sigma \in \Gamma_{p_i}, d(q_i, \sigma q_i) > s} d(q_i, \sigma q_i) e^{-\delta \cdot d(q_i, \sigma q_i)} \to 0$$

as $s \to \infty$. Therefore by Theorem 5.19,

$$e^{-\delta T} \cdot \mu_T(\mathrm{cusp}(r)K) = o_r(1).$$

If we denote by Φ_r a continuous approximation of the unit neighborhood of $\mathcal{C}(\Gamma) - (\cup \mathcal{H}_i(r))$ (that is, $\Phi_r = 1$ on the neighborhood and 0 outside a slightly bigger neighborhood) then Theorem 5.16 implies that

$$e^{-\delta T} \delta \mu_T(f \cdot \Phi_r \otimes \xi) = m^{\mathrm{BMS}}(f \cdot \Phi_r \otimes \xi) + o_T(1).$$

Hence

$$\left| e^{-\delta T} \delta \mu_T(f \otimes \xi) - m^{\mathrm{BMS}}(f \otimes \xi) \right| = o_T(1) + o_r(1) + m^{\mathrm{BMS}}(\mathrm{cusp}(r)K)$$

since the support of m^{BMS} is contained in $\mathcal{C}(\Gamma)$. By taking $r \to \infty$, we finish the proof of the first claim (5.2). In view of (5.20) and Theorem 5.16, the claim on the error term follows as well.

We now deduce (5.3) from (5.2) ; this is done in [18, Section 5]; we recall the proof for the convenience of the reader.

Without loss of generality we may assume $f \otimes \xi \geq 0$. It follows from the definition that

$$\delta T e^{-\delta T} \eta_T(f \otimes \xi) \geq \delta e^{-\delta T} \mu_T(f \otimes \xi). \tag{5.21}$$

Therefore (5.2) implies that

$$\liminf_T \delta T e^{-\delta T} \eta_T(f \otimes \xi) \geq m^{\mathrm{BMS}}(f \otimes \xi).$$

We now bound $\eta(f \otimes \xi)$ from above. Let $\varepsilon > 0$ be small fixed number. We have

$$\delta T e^{-\delta T} \eta_T(f \otimes \xi) = \tag{5.22}$$
$$\delta T e^{-\delta T} \left(\sum_{\mathcal{G}_\Gamma((1-\varepsilon)T)} \mathcal{D}_C(f)\xi(h_C) + \sum_{\mathcal{G}_\Gamma(T)-\mathcal{G}_\Gamma((1-\varepsilon)T)} \mathcal{D}_C(f)\xi(h_C) \right) \leq$$
$$\delta T e^{-\delta T} \left(\sum_{\mathcal{G}_\Gamma((1-\varepsilon)T)} \ell(C)\mathcal{D}_C(f)\xi(h_C) + \sum_{\mathcal{G}_\Gamma(T)-\mathcal{G}_\Gamma((1-\varepsilon)T)} \tfrac{\ell(C)}{(1-\varepsilon)T}\mathcal{D}_C(f)\xi(h_C) \right) \leq$$
$$Te^{-\delta\varepsilon T} \left(\delta e^{-\delta((1-\varepsilon)T)} \mu_{(1-\varepsilon)T}(f \otimes \xi) \right) + \frac{\delta e^{-\delta T}}{1-\varepsilon} \left(\mu_T(f \otimes \xi) - \mu_{(1-\varepsilon)T}(f \otimes \xi) \right).$$

Therefore again by Theorem 5.1,

$$\begin{aligned}\limsup_T \delta T e^{-\delta T} \eta_T(f\otimes\xi) &\leq m^{\mathrm{BMS}}(f\otimes\xi) \left(\limsup_T Te^{-\delta\varepsilon T} + \frac{1}{1-\varepsilon} + e^{-\varepsilon\delta T} \right) \\ &\leq \frac{1}{1-\varepsilon} m^{\mathrm{BMS}}(f \otimes \xi)\end{aligned}$$

Since $\varepsilon > 0$ is arbitrary, this proves the claim. When (5.2) is effective, the above argument also gives an effective statement (instead of using lim sup and lim inf, we compare the error term and ε and take ε appropriately).

References

[1] M. Babillot. *On the mixing property for hyperbolic systems.* Israel J. Math., 129:61–76, 2002.

[2] J. Bourgain, A. Gamburd, and P. Sarnak. *Generalization of Selberg's* 3/16 *theorem and Affine sieve.* Acta Math. 207, (2011) 255-290.

[3] J. Bourgain, A. Kontorovich, and P. Sarnak. Sector estimates for Hyperbolic isometries. GAFA. 20, 1175–1200, 2010.

[4] K. Corlette and A. Iozzi *Limit sets of discrete groups of isometries of exotic hyperbolic spaces* Transaction of the AMS., Vol 351, 1507–1530, 1999

[5] F. Dal'bo, J.-P. Otal, M. Peign. *Séries de Poincar e des groupes géométriquement finis* Israel J. Math. 118 (2000), 109–124.

[6] B. H. Bowditch. *Geometrical finiteness with variable negative curvature.* Duke Math. J., Vol 77, 229-274, 1995

[7] L. Flaminio and R. Spatzier. *Geometrically finite groups, Patterson-Sullivan measures and Ratner's theorem,* Inventiones, **99**, (1990), 601-626.

[8] R. Gangolii and G. Warner. *Zeta functions of Selberg's type for some noncompact quotients of symmetric spaces of rank one,* Nagoya Math. J., **78**, (1980), 1-44.

[9] A. Gorodnik and H. Oh. *Orbits of discrete subgroups on a symmetric space and the Furtstenberg boundary.* Duke Math., Vol 139 (2007), 483–525.

[10] A. Gorodnik, H. Oh, N. Shah. *Strong Wavefront lemma and counting lattice points in sectors.* Israel J. Math., Vol 176 (2010), 419–444.

[11] G. Margulis. *On some aspects of the theory of Anosov systems,* Springer Monographs in Mathematics. Springer-Verlag, Berlin, 2004.

[12] A. Mohammadi and H. Oh. *Matrix coefficients, Counting and Primes for orbits of geometrically finite groups,* J. European Math. Soc., to appear, arXiv:1208.4139

[13] F. Naud *Expanding maps on cantor sets and analytic continuation of zeta functions,* Annales scientifiques de l'École normale suprieure, Vol 38 (2005), 116-153
[14] H. Oh, N. Shah. *Equidistribution and Counting for orbits of geometrically finite hyperbolic groups.* Journal of AMS, Vol 26 (2013), 511–562.
[15] W. Parry and M. Pollicott. *The Chebotarov theorem for Galois coverings of Axiom A flows.* ETDS Vol 6 (1986) 133-148
[16] S. Patterson. *The limit set of a Fuchsian group.* Acta Mathematica, 136:241–273, 1976.
[17] G. Prasad and A. Rapinchuk. *Existence of irreducible $\mathbb{R}$-regular elements in Zariski-dense subgroups.* Mathematical research letters, (10), 21–32, 2003.
[18] T. Roblin. *Ergodicité et équidistribution en courbure négative.* Mém. Soc. Math. Fr. (N.S.), (95):vi+96, 2003.
[19] P. Sarnak. *The arithmetic and geometry of some hyperbolic three manifolds.* Acta Math. 151 (1983), no. 3-4, 253295.
[20] P. Sarnak and M. Wakayama. *Equidistribution of holonomy about closed geodesics.* Duke Math. Journal., Vol 100 (1999)1–57
[21] D. Sullivan. *The density at infinity of a discrete group of hyperbolic motions.* Inst. Hautes Études Sci. Publ. Math., (50):171–202, 1979.
[22] D. Sullivan. *Entropy, Hausdorff measures old and new, and limit sets of geometrically finite Kleinian groups.* Acta Math., 153(3-4):259-277, 1984.
[23] I. Vinogradov. *Effective bisector estimate with applications to Apollonian circle packings.* Preprint, arXiv:1204.5498.
[24] D. Winter. *Mixing of frame flow for rank one locally symmetric spaces and measure classification*, Preprint, arXiv:1403.2425.

Mathematics department, Yale university, New Haven, CT 06511
E-mail address: gregorii.margulis@yale.edu

Department of Mathematics, The University of Texas at Austin, Austin, TX 78750
E-mail address: amir@math.utexas.edu

Mathematics department, Yale university, New Haven, CT 06511 and Korea Institute for Advanced Study, Seoul, Korea
E-mail address: hee.oh@yale.edu