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 $\Gamma$ 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 $G = \text{PSL}_2(\mathbb{C})$, our results imply the equidistribution of eigenvalues of elements of $\Gamma$ in the complex plane.

When $\Gamma$ 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 $\Gamma$ 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 $T^1(X)$ of $X$ can be identified with $\Gamma \backslash G / M$. Each closed geodesic $C$ on $T^1(X)$ 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 $\Gamma$ 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 $\Gamma$ is called geometrically finite if the unit neighborhood of its convex core in $X$ is of finite Riemannian volume (cf. [Bow95]). 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 $G = \text{SO}(n,1)^o$ is

Keywords and phrases: Geometrically finite groups, Closed geodesics, Holonomy, Bowen–Margulis–Sullivan measure
Mathematics Subject Classification: Primary 11N45, 37F35, 22E40; Secondary 37A17, 20F67

Margulis was supported in part by NSF Grant #1265695. Mohammadi was supported in part by NSF Grant #1200388 and Sloan fellowship. Oh was supported in part by NSF Grant #1068094 and #1361673.
the group of its orientation preserving isometries, any discrete group \( \Gamma \) 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 \( \Gamma \). It is well-known that \( \delta > 0 \) if \( \Gamma \) is non-elementary.

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

\[
G_\Gamma(T) := \{ C : C \text{ is a closed geodesic in } T^1(X), \ell(C) \leq T \}.
\]

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

**Theorem 1.1.** Let \( \Gamma \) be geometrically finite and Zariski dense in \( G \). Then for any continuous class function \( \varphi \) on \( M \),

\[
\sum_{C \in G_\Gamma(T)} \varphi(h_C) \sim \frac{e^{\delta T}}{\delta T} \int_M \varphi \, dm \quad \text{as } T \to \infty
\]

where \( dm \) is the Haar probability measure on \( M \).

The asymptotic of \( \# G_\Gamma(T) \) was well-known, due to Margulis [Mar04] for \( X \) compact, to Gangolli and Warner [GW80] for \( X \) noncompact but of finite volume, and to Roblin [Rob03] for \( X \) geometrically finite:

\[
\# G_\Gamma(T) \sim \frac{e^{\delta T}}{\delta T}.
\]

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

If we define \( G_\Gamma^1(T) \) to be the set of all primitive and non-primitive closed geodesics of length at most \( T \), then it is easy to see that

\[
\# G_\Gamma^1(T) = \# G_\Gamma(T) + O(T) \cdot \# G_\Gamma(T/2).
\]

Therefore Theorem 1.1 remains the same if we replace \( G_\Gamma(T) \) by \( G_\Gamma^1(T) \). It is worth mentioning that if one considers all geodesics, then it follows from the work of Prasad and Rapinchuk [PR03] that the set of all holonomy classes about closed geodesics in \( T^1(X) \) is dense in the space of all conjugacy classes of \( M \).

When \( G = \text{SO}(n,1)^0 \) and \( \Gamma \) is a co-compact lattice, Theorem 1.1 was known due to Parry and Pollicott [PP86], who showed that the topological mixing of the frame flow on a compact manifold implies the equidistribution of holonomies. When \( \Gamma \) is a lattice in a general rank one group \( G \), Theorem 1.1 was proved by Sarnak and Wakayama [SW99]; their method is based on the Selberg trace formula and produces an error term. Therefore Theorem 1.1 is new only when \( \Gamma \) 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 $\Gamma$ be a geometrically finite subgroup of $G = \text{SO}(n,1)^\circ$ with $n \geq 3$. We suppose that $\delta > \max\{n-2,(n-2+\kappa)/2\}$ where $\kappa$ is the maximum rank of all parabolic fixed points of $\Gamma$. Then there exists $\eta > 0$ such that for any smooth class function $\varphi$ on $M$,

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

where the implied constant depends only on the Sobolev norm of $\varphi$.

Theorem 1.2 gives a quantitative counting result for closed geodesics of length at most $T$. This was known when $\Gamma$ is a lattice by the work of Selberg, and Gangolli-Warner [GW80] by the trace formula approach, or when $\Gamma$ is a convex co-compact subgroup of $\text{SO}(2,1)$ by Naud [Nau05] 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 4.2 holds; this will be evident from our proof.

As is well-known, the set of closed geodesics in $T^1(X)$ is in one-to-one correspondence with the set of conjugacy classes of primitive hyperbolic elements of $\Gamma$. If $A = \{a_t\}$ denotes the one parameter subgroup whose right translation action on $\Gamma\backslash G/M$ corresponds to the geodesic flow on $T^1(X)$, 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 $\gamma$ in $\Gamma$ and by $[\Gamma_{ph}]$ the set of all conjugacy classes of primitive hyperbolic elements of $\Gamma$. Given a closed geodesic $C$ in $T^1(X)$, 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 $\Gamma$.

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

The aforementioned result of Prasad and Rapinchuk says that any Zariski dense subgroup $\Gamma$ contains a hyperbolic element $\gamma$ such that the argument of the complex number $\lambda_\gamma$ is an irrational multiple of $\pi$ [PR03]. We show a stronger theorem that the arguments of $\lambda_\gamma$’s are equidistributed in all directions when $\Gamma$ is geometrically finite.

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

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

$$\#\{[\gamma] \in [\Gamma_{ph}] : |\lambda_\gamma| < T, \theta_1 < \text{Arg}(\lambda_\gamma) < \theta_2\} \sim \frac{(\theta_2 - \theta_1)T^{2\delta}}{2\pi \delta \log T}$$

as $T \to \infty$. \hfill (1.1)

If $\delta > 1$ and $\Gamma$ has no rank 2 cusp, or if $\delta > 3/2$ in general, then (1.1) 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 $\lambda_\gamma$ is precisely the holonomy associated to $\gamma$. Hence Theorem 1.3 is a special case of Theorems 1.1 and 1.2.

In the case when $\Gamma$ is contained in an arithmetic subgroup of $\text{PSL}_2(\mathbb{C})$, the polynomial error term can be taken to be uniform over all congruence subgroups of $\Gamma$; this follows from our approach based on the work of Bourgain, Gamburd and Sarnak [BGS11] and of Mohammadi and Oh [MO14]. In the case when $\Gamma \subset \text{PSL}_2(\mathbb{O}_D)$ where $\mathbb{O}_D$ is the ring of integers of an imaginary quadratic extension $\mathbb{Q}(\sqrt{-D})$ of $\mathbb{Q}$, the eigenvalues of $\Gamma$ are fundamental units of $\mathbb{O}_D$ (cf. [Sar83]), in which case Theorem 1.3 also bears an arithmetic application on the distribution of such fundamental units arising from $\Gamma$.

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

$$\eta_T(f \otimes \xi) := \sum_{C \in G(T)} D_C(f)\xi(h_C)$$

where $D_C$ 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 $\xi$,

$$\eta_T(f \otimes \xi) \sim \frac{e^{\delta T} \cdot m_{\text{BMS}}(f) \cdot \int M \xi dm}{\delta \cdot |m_{\text{BMS}}| \cdot T} \quad \text{as} \quad T \to \infty$$

where $m_{\text{BMS}}$ is the Bowen–Margulis–Sullivan measure on $T^1(X)$.

We will deduce (1.3) from the following:

$$\mu_T(f \otimes \xi) \sim \frac{e^{\delta T} \cdot m_{\text{BMS}}(f) \cdot \int M \xi dm}{\delta \cdot |m_{\text{BMS}}|} \quad \text{as} \quad T \to \infty$$

where $\mu_T(f \otimes \xi) = \sum_{C \in G(T)} \mathcal{L}_C(f)\xi(h_C)$ for $\mathcal{L}_C = \ell(C) \cdot D_C$ (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 (1.4), the following $\epsilon$-flow boxes play an important role: for $g_0 \in G$, set

$$\mathfrak{B}(g_0, \epsilon) = g_0(N^+_\epsilon N^- \cap N^-N^+ AM)M_\epsilon A_\epsilon.$$

where $A_\epsilon$ (resp. $M_\epsilon$) is the $\epsilon$-neighborhood of $e$ in $A$ (resp. $M$) and $N^\pm_\epsilon$ denotes the $\epsilon$-neighborhood of $e$ in $N^\pm$. Let $\mathfrak{B}(g_0, \epsilon)$ denote the image of $\mathfrak{B}(g_0, \epsilon)$ under the canonical projection $G \to \Gamma \backslash G/M$. Fixing a Borel subset $\Omega$ of $M$ which is conjugation-invariant, the main idea is to relate the restriction of $\mu_T$ to $\mathfrak{B}(g_0, \epsilon) \otimes \Omega$ with the counting function of the set $\Gamma \cap \mathfrak{B}(g_0, \epsilon)A_T^{-1}\Omega$ with $A_T^+ = \{a_t : 0 < t \leq T\}$ (see comparison Lemma 5.5); 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 (1.5). This
flow box was first used in Margulis’ work on counting closed geodesics [Mar04]. The
counting function of $\Gamma \cap \mathcal{B}(g_0, \varepsilon) A^+_T \mathcal{B}(g_0, \varepsilon)^{-1}$ can then be understood based on the
mixing result of Winter [Win14], which says that the $A$ action on $L^2(\Gamma \backslash G, m^{BMS})$
is mixing. An effective mixing statement for the cases mentioned in Theorem 1.2
was obtained in [MO14]. We also remark that if we restrict ourselves only to those
$f$ with compact support, then (1.4) holds for any discrete subgroup $\Gamma$ admitting a
finite BMS measure; that is, $\Gamma$ need not be geometrically finite.

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: $\text{SO}(n, 1)^\circ$, $\text{SU}(n, 1)$,
$\text{Sp}(n, 1)$ ($n \geq 2$) and $F_{-20}$, 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

$$D(\tilde{X}) = \lim_{T \to \infty} \log \text{Vol}(B(o, T)) / T$$

(2.1)

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

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

(2.2)

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

We denote by $\partial_\infty(\tilde{X})$ the geometric boundary of $\tilde{X}$ and by $T^1(\tilde{X})$ the unit
tangent bundle of $\tilde{X}$. Fixing a vector $v_o \in T^1(\tilde{X})$ based at $o$, $T^1(\tilde{X})$ can be identified
with $G/M$ where $M$ is the stabilizer of $v_o$ in $G$. For a vector $v \in T^1(\tilde{X})$, 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 $T^1(\tilde{X})$; 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 \( \partial_\infty(\tilde{\mathcal{X}}) \).

Let \( d = d_G \) be a left \( G \)-invariant Riemannian metric on \( G \) which induces the metric \( d_{\tilde{\mathcal{X}}} \) on \( \tilde{\mathcal{X}} = G/K \). 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 \( \varepsilon \)-ball at \( g_0 \) with \( S \). Hence the \( \varepsilon \)-balls \( G_\varepsilon(g_0) \) form a basis of open neighborhoods at \( g_0 \).

**Flow box:** Following Margulis [Mar04], we will define the flow-box around \( g_0 \in G \) for all small \( \varepsilon > 0 \). For this, we will use the following \( \varepsilon \)-neighborhoods of \( e \) in \( N^+, N^-, A, M \).

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

\[
N^\pm_\varepsilon := \{ n^\pm_\varepsilon := \exp x \in N^\pm : \|x\| < \varepsilon \}
\]

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

For \( A \) and \( M \), we simply put

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

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

\[
\mathfrak{B}(g_0, \varepsilon) = g_0(N^+_\varepsilon N^- \cap N^+_\varepsilon N^+ AM)M_\varepsilon A_\varepsilon.
\]

(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 2.1 (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 \( \{ t \in \mathbb{R} : g a_t \in \mathfrak{B}(g_0, \varepsilon) \} \) is of Lebesgue length \( 2\varepsilon \);
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

\[
G_{c^{-1}\varepsilon}(g_0) \subset \mathfrak{B}(g_0, \varepsilon) \subset G_{c\varepsilon}(g_0);
\]

(2.4)

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 \( \text{Fix}(g) = \{ x \in \tilde{X} \cup \partial(\tilde{X}) : g(x) = x \} \) 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 2.2.** 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 \( a_1 = a_2 \), \( m_1 = mm_2m_2^{-1} \) for some \( m \in M \) and \( h \in AM \).

**Proof.** For \( g \in G \), define
\[
N^\pm(g) = \{ q \in G : g^\ell qg^{-\ell} \to e \text{ as } \ell \to \pm \infty \}.
\]
Putting \( g_i = a_im_i \in A^+ M \) for \( i = 1, 2 \), we have \( N^\pm(g_i) = N^\pm \). On the other hand, since \( g_2 = h_1g_1h^{-1} \), 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 \( ha_1m_1h^{-1} = a_1(mm_1m_2^{-1}) = a_2m_2 \).

Hence \( a_2^{-1}a_1 \in A \cap M = \{ e \} \); so \( a_1 = a_2 \), as well as \( m_2 = mm_1m_2^{-1} \).

As an immediate corollary, we have:

**Corollary 2.3.** If a hyperbolic element \( g \in G \) is of the form:
\[
g = h_2a_gmgh_g^{-1}
\]
with \( a_gm_g \in A^+ M \), then \( a_g \) is uniquely determined, \( m_g \in M \) is determined unique up to conjugation and \( R_g := h_gAe_0 \) is independent of the choice of \( h_g \).

The geodesic \( R_g := h_gAe_0 \subset \tilde{X} \) is called the oriented axis of \( g \); \( g \) preserves \( R_g \), and acts as a translation by \( T := d(a_g, e) \).

Let \( \Gamma \) be a torsion-free and non-elementary discrete subgroup of \( G \). A closed geodesic \( C \) of length \( T > 0 \) on \( T^1(X) = \Gamma\backslash G/M \) is a compact set of the form \( \Gamma\backslash gAM/M \) for some \( g \in G \) such that \( gAMg^{-1} \cap \Gamma \) is generated by a hyperbolic element \( \gamma = ga_\gamma m_\gamma g^{-1} \) 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
\[
\Gamma\backslash gM_0aT = \Gamma\backslash gM_0m
\]
for some \( m_0, m \in M \), then \([m] = [m_\gamma]\). Geometrically, \( \Gamma\backslash gM_0 \) is a frame which contains the tangent vector \( \Gamma\backslash gM \), and the element \( m \) measures the extent to which parallel transport around the closed geodesic \( gM_0aT \) differs from the original frame \( \Gamma\backslash gM_0 \). 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.
3 Effective closing lemma

Let $\Gamma$ 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 $\Gamma$. We denote by $\pi : G \rightarrow \Gamma \backslash G$ the canonical projection map.

For two elements $h_1, h_2 \in G$, we will write $h_1 \sim_\varepsilon h_2$ if $d_G(h_1, h_2) \leq \varepsilon$ and $h_1 \sim_{O(\varepsilon)} h_2$ if $d_G(h_1, h_2) \leq c\varepsilon$ 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 $\varepsilon$ flow box $\mathcal{B}(g_0, \varepsilon)$ injects to $\Gamma \backslash G$. In what follows, we will consider boxes $\mathcal{B}(g_0, \varepsilon)$ only for those $\varepsilon$ 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 $\varepsilon$-box $\pi(\mathcal{B}(g_0, \varepsilon))$, that is, there exist $g_1, g_2 \in \mathcal{B}(g_0, \varepsilon)$ 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 [Mar04, Lemma 6.2], [Rob03, Chapter 5]) says that there is a closed geodesic nearby; more precisely, $\gamma$ is a hyperbolic element and its oriented axis $R_\gamma$ is nearby the box $\pi(B(g_0, \varepsilon))$ in the space $\Gamma \setminus G/M = T^1(X)$. We will need more detailed information on this situation. We will show that the oriented axis $R_\gamma$ passes through $O(\varepsilon e^{-T})$-neighborhood of the box $\pi(B(g_0, \varepsilon))$ and $\tilde{a}_\gamma$ and $[\tilde{m}_\gamma]$ are $O(\varepsilon)$-close to $a_\gamma$ and $[m_\gamma]$ respectively where $a_\gamma$ (Fig. 1) and $[m_\gamma]$ are defined as in (2.5) for $\gamma$.

**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 $g_1, g_2 \in B(g_0, \varepsilon)$ and $\gamma \in G$ such that

$$g_1 \tilde{a}_\gamma \tilde{m}_\gamma = \gamma g_2 \quad (3.1)$$

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

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

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 $g_3 := g_0 b_\varepsilon \in B(g_0, \varepsilon)$,

$$g_3^{-1} g_3 = n_w^+ a'_\gamma m'_\gamma n_z^- \quad (3.2)$$

where $a'_\gamma m'_\gamma \in \tilde{a}_\gamma \tilde{m}_\gamma A_{O(\varepsilon)} M_{O(\varepsilon)}$, $n_w^+ \in N_{O(e^{-\tau})}$ and $n_z^- \in N_{O(e^{-\tau})}$.

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

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

we may write $b_\varepsilon$ and $d_\varepsilon$ as follows:

$$b_\varepsilon = b_\varepsilon^+ n_x^- b_\varepsilon^0 \in N^+ N^- (A_\varepsilon M_\varepsilon);$$

$$d_\varepsilon = d_\varepsilon^+ n_y^+ d_\varepsilon^0 \in N^- N^+ (AM).$$

By Lemma 2.1, 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

$$g_0 b_\varepsilon^+ n_x^- \alpha^{(1)} \tilde{m}_\gamma^{(1)} = \gamma g_0 d_\varepsilon^+ n_y^+ \quad (3.3)$$

where $\alpha^{(1)} \tilde{m}_\gamma^{(1)} := b_\varepsilon^0 \tilde{a}_\gamma \tilde{m}_\gamma (d_\varepsilon^0)^{-1} \in AM$. 
By the transversality between $N^{-}$ and $AM N^{+}$, we obtain a unique element $n^{-}_{x} \in N^{-}_{O(\varepsilon)}$ satisfying that

$$b^{+}_{\varepsilon}n^{-}_{x} \in d^{-}_{\varepsilon}(N^{+}_{O(\varepsilon)}A_{O(\varepsilon)}M_{O(\varepsilon)}).$$

(3.4)

Since $b^{+}_{\varepsilon} \in N^{+}_{\varepsilon}$ and $d^{-}_{\varepsilon} \in N^{-}_{\varepsilon}$, we have

$$h_{\varepsilon} := b^{+}_{\varepsilon}n^{-}_{x} \in (N^{+}_{\varepsilon}N^{-} \cap N^{-}_{\varepsilon}N^{+} AM) \subset \mathcal{B}(\varepsilon)$$

and hence

$$g_{0}h_{\varepsilon} \in \mathcal{B}(g_{0}, \varepsilon).$$

Now setting $g_{3} := g_{0}h_{\varepsilon}$, we claim that (3.2) holds. By (3.4), $h_{\varepsilon} = d^{-}_{\varepsilon}n^{-}_{x}a^{-1}_{\varepsilon}m^{-1}_{\varepsilon}a^{-1}_{\varepsilon}m^{-1}_{\varepsilon}m^{-1}_{\varepsilon}n^{-}_{x} \in d^{-}_{\varepsilon}(N^{+}_{O(\varepsilon)}A_{O(\varepsilon)}M_{O(\varepsilon)}).$ Rewriting (3.3), we have

$$g_{0}h_{\varepsilon}(n^{-}_{x})^{-1}n^{-}_{x}a^{-1}_{\gamma}m^{-1}_{\gamma} = \gamma g_{0}h_{\varepsilon}a^{-1}_{\varepsilon}m^{-1}_{\varepsilon}n^{-}_{x} \in N^{-}_{O(\varepsilon)},$$

and hence

$$g_{3}^{-1}\gamma g_{3} = (n^{-}_{x})^{-1}n^{-}_{x}(a^{-1}_{\gamma}m^{-1}_{\gamma}a^{-1}_{\varepsilon}m^{-1}_{\varepsilon}n^{-}_{x})^{-1}(n^{-}_{x})^{-1}n^{-}_{x} \in N^{-}_{O(\varepsilon)},$$

where $a^{-1}_{\gamma}m^{-1}_{\gamma} := (a^{-1}_{\gamma}m^{-1}_{\gamma}a^{-1}_{\varepsilon}m^{-1}_{\varepsilon}n^{-}_{x})^{-1}n^{-}_{x} \in N^{-}_{O(\varepsilon)}$. We claim that $a^{-1}_{\gamma}m^{-1}_{\gamma} \in N^{-}_{O(e-\varepsilon)}$ and we can write

$$n^{-}_{z_{2}}n^{-}_{w_{1}} = n^{-}_{w_{2}}a^{-1}_{\gamma}m^{-1}_{\gamma}n^{-}_{z_{3}} \in N^{+}_{O(\varepsilon)}A_{O(\varepsilon)}M_{O(\varepsilon)}N^{-}_{O(e-\varepsilon)}.$$}

Therefore

$$g_{3}^{-1}\gamma g_{3} = a^{-1}_{\gamma}m^{-1}_{\gamma}n^{-}_{z_{2}}n^{-}_{w_{1}} = n^{-}_{z_{3}}a^{-1}_{\gamma}m^{-1}_{\gamma}n^{-}_{z_{3}}$$

where $a^{-1}_{\gamma}m^{-1}_{\gamma} := (a^{-1}_{\gamma}m^{-1}_{\gamma}a^{-1}_{\varepsilon}m^{-1}_{\varepsilon})^{-1}n^{-}_{x} \in N^{+}_{O(e-\varepsilon)}$. This proves the claim (3.2).

**Step 2:** Set $g := g_{3}^{-1}\gamma g_{3}$ so that $g = n^{-}_{w}a^{-1}_{\gamma}m^{-1}_{\gamma}n^{-}_{x}$. We claim that

$$g \in (n^{-}_{x}n^{-}_{y})a^{-1}_{\gamma}A_{O(\varepsilon)}m^{-1}_{\gamma}M_{O(\varepsilon)}(n^{-}_{x}n^{-}_{y})^{-1}$$

(3.5)

with $n^{-}_{x} \in N^{+}_{O(e-\varepsilon)}$ and $n^{-}_{y} \in N^{-}_{O(e-\varepsilon)}$.

For $n^{-}_{x}$ as above, for any $n^{+}_{x} \in N^{+}_{\varepsilon}$, there exists a unique element $n^{+}_{\alpha(x)} \in N^{+}_{O(\varepsilon)}$ such that $(n^{-}_{x})n^{+}_{x} \in n^{+}_{\alpha(x)}A_{\varepsilon}M_{\varepsilon}N^{-}_{\varepsilon}$. Moreover the map $n^{+}_{x} \mapsto n^{+}_{\alpha(x)}$ is a diffeomorphism of $N^{+}_{\varepsilon}$ 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 \(N^+_{\varepsilon}\) onto its image \(N^+_{\varepsilon+O(\varepsilon^{-\varepsilon})}\). Since \(n^+_w \in N^+_{O(\varepsilon^{-\varepsilon})}\), if \(T\) is large enough, we can find \(n^+_x \in N_{O(\varepsilon^{-\varepsilon})}\) 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 \((n^+_{\alpha(x)})^{-1}n^+_w = a_\varepsilon \varepsilon m_\varepsilon n^+_x(n^+_x)^{-1}\) with \(n^+_w \in N^+_{O(\varepsilon^{-\varepsilon})}\), 
\(a_\varepsilon \in A_{O(\varepsilon)}\) and \(m_\varepsilon \in M_{O(\varepsilon)}\). Therefore, plugging in these, 
\[g = n^+_w a'_\gamma m'_\gamma n^+_w\] 
\[= n^+_x a'_\gamma m'_\gamma (n^+_{\alpha(x)})^{-1}n^+_w\] 
\[= n^+_x (a'_\gamma m'_\gamma) n^+_w (n^+_x)^{-1}\]
where \(a''_\gamma = a'_\gamma a_\varepsilon\) and \(m''_\gamma = m'_\gamma m_\varepsilon\). Since the map 
\[n^-_y \mapsto (a''_\gamma m''_\gamma)^{-1} n^-_y (a''_\gamma m''_\gamma)(n^-_y)^{-1}\]
is a diffeomorphism of \(N^-_{\varepsilon}\) onto its image \(N^-_{\varepsilon+O(\varepsilon^{-\varepsilon})}\), for all large \(T\), we can find 
\(n^-_y \in N^+_{O(\varepsilon^{-\varepsilon})}\) such that 
\[n^-_u = (a''_\gamma m''_\gamma)^{-1}(n^-_y)^{-1}(a''_\gamma m''_\gamma) n^-_u\].
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 \(g_4 := g_3 n^+_x (n^-_y)^{-1} \in \mathfrak{B}(g_0, \varepsilon + O(\varepsilon^{-\varepsilon}))\). 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 3.2** (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 
\[g_1 \tilde{a}_\gamma \tilde{m}_\gamma = \gamma g_2\]
(3.6)
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 \(\gamma\) is primitive, i.e., \(\gamma\) cannot be written as a power of another element of \(\Gamma\). Then there exists an element \(g_\gamma \in \mathfrak{B}(g_0, \varepsilon + O(\varepsilon^{-\varepsilon}))\) such that
4 Counting results for $\Gamma \cap \mathcal{B}(g_0, \varepsilon) A_T \Omega \mathcal{B}(g_0, \varepsilon)^{-1}$

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

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

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

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 $\# \Gamma \cap \mathcal{B}(g_0, \varepsilon) A_T^\pm \Omega \mathcal{B}(g_0, \varepsilon)^{-1}$.

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 $\Gamma$ is Zariski dense and that the Bowen–Margulis–Sullivan measure, the BMS measure for short, on $T^1(X) = \Gamma \backslash G / M$ 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 $\Gamma$, which is the set of all accumulation points in $X \cup \partial(X)$ of an orbit of $\Gamma$ in $X$. Denote by $\delta = \delta_\Gamma$ the critical exponent of $\Gamma$. Denote by $\{\nu_x : x \in \tilde{X}\}$ a $\Gamma$-invariant conformal density of dimension $\delta$ supported on the limit set $\Lambda(\Gamma)$; such a density exists by the construction given by Patterson [Pat76]. 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 $T^1(\tilde{X})$ 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 $T^1(\tilde{X})$ with respect to $\{\nu_x\}$ is defined as follows:

$$d\tilde{m}^{BMS}(u) = \frac{d\nu_x(u^+) d\nu_x(u^-) d\varepsilon}{d_x(u^+, u^-)^{2\delta}}.$$
The definition is independent of \( x \in \tilde{X} \) and \( \tilde{m}^{\text{BMS}} \) is right \( A \)-invariant and left \( \Gamma \)-invariant, and hence induces a geodesic flow invariant Borel measure on \( T^1(X) \), which we denote by \( m^{\text{BMS}} \). If \( |m^{\text{BMS}}| < \infty \), then the geodesic flow is ergodic with respect to \( m^{\text{BMS}} \), as shown by Sullivan [Sul79] and moreover mixing by Babillot [Bab02].

As we are eventually interested in counting a \( \Gamma \) orbit in a family \( \Xi_1 A^+_\Omega \Xi_2 \) with \( \Omega \) any Borel subset in \( M \), we need to understand the mixing phenomenon for the \( A \)-action on \( \Gamma \setminus G \), not only on \( \Gamma \setminus G/M \). By abuse of notation, we denote by \( m^{\text{BMS}} \) the \( M \)-invariant lift of \( m^{\text{BMS}} \) to \( \Gamma \setminus G \). Winter [Win14] showed that if \( \Gamma \) is Zariski dense and \( |m^{\text{BMS}}| < \infty \), then the \( A \)-action on \( \Gamma \setminus G \) is mixing for this extension \( m^{\text{BMS}} \); this was earlier claimed in [FS90] for the case of \( G = \text{SO}(n, 1)^\circ \) and \( \Gamma \) geometrically finite.

In the rest of this section, we assume that \( \Gamma \) is Zariski dense and \( |m^{\text{BMS}}| < \infty \).

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

\[
d\tilde{m}^{\text{BR}}(kma_n^+) = e^{-\delta r}dn^+dsd\nu_o(kv_o^-)dm \quad \text{for} \quad kma_n^+ \in (K/M)MAN^+; \quad (4.1)
\]

\[
d\tilde{m}^{\text{BR}}_* (kma_n^-) = e^{\delta r}dn^-dsd\nu_o(kv_o^+)dm \quad \text{for} \quad kma_n^- \in (K/M)MAN^- \quad (4.2)
\]

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:

\[
dg = d\tilde{m}^{\text{Haar}}(a_nk) = dsdn^\pm dk
\]

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

The following theorem can be deduced from the mixing of \( m^{\text{BMS}} \), as observed first by Roblin for \( M \)-invariant functions ([Rob03], see also [OS13]).

**Theorem 4.1** ([Rob03], [OS13], [Win14]). For any functions \( \Psi_1, \Psi_2 \in C_c(\Gamma \setminus G) \),

\[
\lim_{t \to +\infty}e^{(D-\delta)t} \int_{\Gamma \setminus G} \Psi_1(gat)\Psi_2(g)dg = \frac{m^{\text{BR}}(\Psi_1) \cdot m^{\text{BR}}_*(\Psi_2)}{|m^{\text{BMS}}|}
\]

where \( D = D(\tilde{X}) \) is the volume entropy of \( \tilde{X} = G/K \) (see (2.1) and (2.2)).
The quotient by $\Gamma$ of the convex hull of $\Lambda(\Gamma)$ is called the convex core of $\Gamma$. A discrete group $\Gamma$ is called \textit{geometrically finite} if the volume of a unit neighborhood of the convex core of $\Gamma$ is finite. Clearly lattices are geometrically finite. If $\Gamma$ is geometrically finite, then $m_{\text{BMS}}$ is known to be finite and the critical exponent is known to be equal to the Hausdorff dimension of $\Lambda(\Gamma)$ ([Sul84,CI99]).

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 4.2** [MO14]. Suppose that $\Gamma$ is a geometrically finite subgroup of $\text{SO}(n, 1)^{\circ}$ 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 $\Psi_1, \Psi_2 \in C_\infty^0(\Gamma \setminus G)$, as $t \to +\infty$,

$$e^{(n-1-\delta)t} \int_{\Gamma \setminus G} \Psi_1(g a_t) \Psi_2(g) dg = \frac{m_{\text{BR}}(\Psi_1) \cdot m_{\text{BMS}}(\Psi_2)}{|m_{\text{BMS}}|} + O(e^{-\varepsilon_0 t})$$

where the implied constant depends only on the Sobolev norms of $\Psi_1$ and $\Psi_2$.

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

$$\mathcal{S}_T(\Xi_1, \Xi_2, \Omega) = \Xi_1 A_T^+ \Omega \Xi_2.$$  \hspace{1cm} (4.3)

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

**Theorem 4.3.** Fix $g_0 \in G$. If $\nu_o(\partial(\Xi_1 v_o^+)) = 0 = \nu_o(\partial(\Xi_2 v_o^-))$ and $\text{Vol}(\partial(\Omega)) = 0$, then as $T \to \infty$,

$$\# \Gamma \cap g_0 \mathcal{S}_T(\Xi_1, \Xi_2, \Omega) g_0^{-1} \sim \frac{\nu_o(\partial(\Xi_1 v_o^+)) \nu_o(\partial(\Xi_2 v_o^-)) \text{Vol}(\Omega)}{\delta|m_{\text{BMS}}|} e^{\delta T}.$$  

Under the assumption of Theorem 4.2, we will prove an effective version of Theorem 4.3. As usual, in order to state a result which is effective, we need to assume certain regularity condition on the boundaries of the sets $\Xi_1, \Xi_2, \Omega$ involved.

**Definition 4.4.** A Borel subset $\Theta \subset \partial(\tilde{X})$ is called admissible with respect to $\nu_o$ if there exists $r > 0$ such that for all small $\rho > 0$,

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

**Remark 4.5.** In the group $G = \text{SO}(2, 1)^{\circ}$, the boundary $\partial(\tilde{X})$ is a circle, and any interval of $\partial(\tilde{X})$ is admissible. For $G = \text{SO}(n, 1)^{\circ}$ with $n \geq 3$, if $\delta > \max\{n - 2, (n - 2 + \kappa)/2\}$ where $\kappa$ is the maximum rank of parabolic fixed points of $\Gamma$, then any Borel subset $\omega$ 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 [MO14], using Sullivan’s shadow lemma.
Theorem 4.6. Let $G$ and $\Gamma$ be as in Theorem 4.2. Suppose that $\Xi_1 v_0^+$ and $\Xi_2^{-1} v_0^-$ 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 $T \to \infty$,

\[
\# \Gamma \cap g_0 S_T(\Xi_1, \Xi_2, \Omega) g_0^{-1} = \left( \frac{\nu_{g_0}(a) g_0 \Xi_1 v_0^+ \nu_{g_0}(a) g_0 \Xi_2^{-1} v_0^-}{\delta |m^{BMS}|} \right) \text{Vol}(\Omega) e^{\delta T} + O(e^{(\delta - \varepsilon_0)T}).
\]

The rest of this section is devoted to the proof of Theorems 4.3 and 4.6. In the case when $G = \text{SO}(n, 1)^\circ$, an analogous theorem for bisectors in $\text{K}A^+ K$ was proved in [MO14] (see also [BKS10] for $n = 2$, [Vin14] for $n = 3$ when $\delta$ is big and [GO07] when $\Gamma$ is a lattice). In view of Theorem 4.1 for a general rank one homogeneous space admitting a finite BMS measure, the proof of Theorem 4.3 is very similar to the one given in [MO14] in principle.

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

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

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

\[
\langle F_B, \Psi_1 \otimes \Psi_2 \rangle_{\Gamma \setminus G} := \int_{\Gamma \setminus G} F_B(g_1; g_2) \Psi_1(g_1) \Psi_2(g_2) dg_1 dg_2.
\]

By a standard folding and unfolding argument, we have

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

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

We deduce that

\[
\langle F_B, \psi^\varepsilon \otimes \Psi^\varepsilon \rangle_{\Gamma \setminus G} = \int_{x \in B} \int_{\Gamma \setminus G} \psi^\varepsilon(g) \Psi^\varepsilon(gx) dg dx
\]

writing $x = n_1 a_t m n_2 \in N^+ AMN^-$ and using $dx = e^{Dt} dn_1 dt dmdn_2$

\[
= \int_{n_1 a_t m n_2 \in B} \left( \int_{\Gamma \setminus G} \psi^\varepsilon(gn_1 a_t m n_2) dg \right) e^{Dt} dt dn_1 dmdn_2
\]

\[
= \int_{n_1 a_t m n_2 \in B} \left( \int_{\Gamma \setminus G} \psi^\varepsilon(gn_1^{-1}) \psi^\varepsilon(ga_t m n_2) dg \right) e^{Dt} dt dn_1 dmdn_2
\]
by applying Theorem 4.1
\[
= \int_{n_1 a_x mn_2 \in B} e^{\delta t} (1 + o(1)) \overline{m}^{BR} (n_1^{-1} \psi \epsilon) m^{BR} ((mn_2) \psi \epsilon) dtn_1 dm n_2
\]
\[
= \int_{n_1 a_x mn_2 \in B} e^{\delta t} (1 + o(1)) \overline{m}^{BR} (n_1^{-1} \psi \epsilon) \overline{m}^{BR} (mn_2 \psi \epsilon) dtn_1 dm n_2. \tag{4.4}
\]

If we define a function \( f_B \) on \( N^+ \times MN^- \) by
\[
 f_B(n_1, mn_2) = \int_{a_x \in n_1^{-1} Bn_2^{-1} m_1 \cap A^+} e^{\delta t} dt,
\]
and a function on \( G \times G \) by
\[
((\psi \epsilon \otimes \psi \epsilon) * f_B)(g, h)
\]
\[
= \int_{n_1 mn_2 \in N^+ MN^-} \psi \epsilon (gn_1^{-1}) \psi \epsilon (hn_2) f_B(n_1, mn_2) dm n_1 dm n_2
\]
\[
= \int_{n_1 mn_2 \in N^+ MN^-} \psi \epsilon (gn_1) \psi \epsilon (hn_2) f_B(n_1^{-1}, mn_2) dm n_1 dm n_2,
\]
then we may write
\[
\langle F_B, \Psi \epsilon \otimes \Psi \epsilon \rangle_{\Gamma \setminus G \times \Gamma \setminus G} = (\overline{m}^{BR} \otimes \overline{m}^{BR}) ((\psi \epsilon \otimes \psi \epsilon) * f_B) + o \left( \max_{n_1 a_x mn_2 \in B} e^{\delta t} \right). \tag{4.5}
\]

Observe that
\[
(\overline{m}^{BR} \otimes \overline{m}^{BR}) ((\psi \epsilon \otimes \psi \epsilon) * f_B)
\]
\[
= \int_{N^+ MN^-} f_B(n_1^{-1}, mn_2) \left( \int_{G \times G} \psi \epsilon (g_1 n_1) \psi \epsilon (h_1 mn_2) d\overline{m}^{BR} (g_1) d\overline{m}^{BR} (h_1) \right) dm n_1 dm n_2. \tag{4.6}
\]

By (4.1) and (4.2), we have
\[
d\overline{m}^{BR} (g_1) d\overline{m}^{BR} (h_1) = e^{\delta (r - r_0)} d\nu_o (k_1 v_o^+) d\nu_o (k_0 v_o^-) \tag{4.7}
\]
for \( g_1 = kn_1 a_x n \in (K/M) MAN^- \) and \( h_1 = k_0 m_1 a_x n_0 n_0 \in (K/M) MAN^+ \).

For \( x \in G \), let \( n_1 (x) \) be the \( N^+ \) component of \( x \) in \( MAN^- N^+ \) decomposition and \( n_2 (x) \) 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 (4.6), first change the inner integral using (4.7) and then perform the change of variables by putting \( g = m_1 a_x n_1 \in MAN^- N^+ \) and
we obtain
\[(m^*_{\text{BR}} \otimes \tilde{m}^*_{\text{BR}})((\psi^\varepsilon \otimes \tilde{\psi}^\varepsilon) * f_B)\]
\[= \int_{k \in K/M, k_0 \in K/M, m_0 \in M} \int_{G \times G} \psi^\varepsilon(kg)\psi^\varepsilon(k_0m_0h)f_B(n_1(g) - 1, n_2(h))e^{\varepsilon(I_1(g) - I_2(h))} \]
\[= \int_{k \in K/M, k_0, k_0' \in K/M} \int_{G \times G} \psi^\varepsilon(g)\psi^\varepsilon(h)f_B(n_1(k^{-1}g) - 1, n_2(k_0^{-1}h))e^{\varepsilon(I_1(k^{-1}g) - I_2(k_0^{-1}h))} \]
d\gamma d\nu_0(k_0) dm_0
\[= \int_{k \in K/M, k_0, k_0' \in K/M} \int_{G \times G} \psi^\varepsilon(g)\psi^\varepsilon(h)f_B(n_1(k^{-1}g) - 1, n_2(k_0^{-1}h))e^{\varepsilon(I_1(k^{-1}g) - I_2(k_0^{-1}h))} \]
d\gamma d\nu_0(k_0) dm_0

where \(d\nu_0(k_0) := d\nu_0(k_0'v_o^{-}) dm\) for \(k_0 = k_0' \times m \in K/M \times M\).

In order to prove Theorem 4.3, we now put
\[S_T := S_T(\tilde{\Xi}_1, \tilde{\Xi}_2, \Omega) \text{ and } F_T := F_{S_T}.\]

Observe that
\[F_T(e, e) = \#(\Gamma \cap S_T(\Xi_1, \Xi_2, \Omega)).\]

Let
\[S^+_{T, \varepsilon} = \cup_{g_1, g_2 \in G_\varepsilon(e)} g_1 S_T g_2 \text{ and } S^-_{T, \varepsilon} = \cap_{g_1, g_2 \in G_\varepsilon(e)} g_1 S_T g_2.\]

We then have
\[\langle F_{S^-_{T, \varepsilon}}, \Psi^\varepsilon \otimes \Psi^\varepsilon \rangle \leq F_T(e, e) \leq \langle F_{S^+_{T, \varepsilon}}, \Psi^\varepsilon \otimes \Psi^\varepsilon \rangle.\]

Together with the strong wave front property for the \(AN^\pm K\) decompositions [GOS10], (4.8) with \(B = S^\pm_{T, \varepsilon}\) now implies that
\[(m^*_{\text{BR}} \otimes \tilde{m}^*_{\text{BR}})((\psi^\varepsilon \otimes \tilde{\psi}^\varepsilon) * f_{S^\pm_{T, \varepsilon}})\]
\[= (1 + O(\varepsilon')) \int_{K/M \times K} f_{S_T}(n_1(k^{-1} - 1, n_2(m^{-1}k_0^{-1})) d\nu_0(k_0^{-1}v_o^{-}) d\nu_0(k_0)\]
\[= (1 + O(\varepsilon')) \frac{e^{\varepsilon T}}{\delta} \nu_0(\Xi_1 v_o^+) \nu_0(\Xi_2^{-1} v_o^-) \text{Vol}(\Omega)\]

where \(\varepsilon' > 0\) goes to 0 as \(\varepsilon \to 0\). Hence, (4.9), (4.4) and (4.5) yield that
\[F_T(e, e) = (1 + O(\varepsilon')) \frac{e^{\varepsilon T}}{\delta} \nu_0(\Xi_1 v_o^+) \nu_0(\Xi_2^{-1} v_o^-) \text{Vol}(\Omega) + o(e^{\varepsilon T}).\]

Since \(\varepsilon > 0\) is arbitrary, we have
\[F_T(e, e) \sim \frac{e^{\varepsilon T}}{\delta} \nu_0(\Xi_1 v_o^+) \nu_0(\Xi_2^{-1} v_o^-) \text{Vol}(\Omega).\]
In order to prove Theorem 4.6 for $g_0 = e$, we note that $O(e^{\delta T})$ in (4.4) can be upgraded into $O(e^{(\delta - \epsilon_0) T})$ in view of Theorem 4.2, and that $O(e')$ in (4.10) can be taken as $O(e^q)$ for some fixed $q > 0$ (we refer to [MO14] for details).

Therefore we get

$$F_T(e, e) = (1 + O(e^q)) \frac{e^{\delta T}}{\delta} \nu_o(\Xi_1 v_o^+ \nu_o(\Xi_2^{-1} v_o^-) \text{Vol}(\Omega) + O(e^{(\delta - \epsilon_0) T}).$$

By taking $\epsilon$ so that $e^q = e^{-\epsilon_0}$, we then obtain

$$F_T(e, e) = \frac{e^{\delta T}}{\delta} \nu_o(\Xi_1 v_o^+ \nu_o(\Xi_2^{-1} v_o^-) \text{Vol}(\Omega) + O(e^{(\delta - \epsilon_1) T})$$

for some positive $\epsilon_1 > 0$. This proves Theorems 4.3 and 4.6 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

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

Moreover, if we set $\nu_{T_0} := g_0^* \nu_g(x)$, then $\{\nu_{T_0} : x \in X \}$ is a $\Gamma_0$-invariant conformal density of dimension $\delta = \delta_{\Gamma_0}$, and the corresponding BMS-measure $m_{\Gamma_0}^{\text{BMS}}$ with respect to $\{\nu_{T_0} \}$ has the same total mass as $m^{\text{BMS}}$. Therefore

$$\frac{\nu_{T_0, o}(\Xi_1 v_o^+) \nu_{T_0, o}(\Xi_2^{-1} v_o^-)}{\delta_{\Gamma_0} |m_{\Gamma_0}^{\text{BMS}}|} = \frac{\nu_{g_0}(g_0 \Xi_1 v_o^+) \nu_{g_0}(g_0 \Xi_2^{-1} v_o^-)}{\delta |m^{\text{BMS}}|}.$$

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

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

$$B(g_0, \epsilon) = g_0 (N^+_\epsilon N^- \cap N^- \cap N^+ \cap AM) M_{\epsilon} A_{\epsilon}.$$  (4.11)

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

$$\tilde{B}(g_0, \epsilon) = \tilde{\pi}(B(g_0, \epsilon)).$$  (4.12)

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

$$m^{\text{BMS}}(f \otimes \xi) := \int_{\Gamma \setminus G / M} f dm^{\text{BMS}} \cdot \int_M \xi dm.$$

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

**Lemma 4.7.** For all small $\epsilon > 0$,

$$m^{\text{BMS}}(\tilde{B}(g_0, \epsilon) \otimes \Omega) = (1 + O(\epsilon)) 2\epsilon \cdot \nu_{g_0}(g_0 N^+_\epsilon N^+) \nu_{g_0}(g_0 N^-) \nu_{g_0}(g_0 N^-) \text{Vol}(\Omega)$$

where the implied constant is independent of $\epsilon > 0$.
Proof. Clearly we have \( m^{\text{BMS}}(\hat{\mathcal{B}}(g_0, \varepsilon) \otimes \Omega) = m^{\text{BMS}}(\mathcal{B}(g_0, \varepsilon) \otimes \Omega) \). Recall that the BMS measure on \( T^1(\hat{X}) \) is given as

\[
\tilde{m}^{\text{BMS}}(u) = \frac{d\nu_{g_0(o)}(u^+)d\nu_{g_0(o)}(u^-)}{d\nu_{g_0(o)}(u^+, u^-)^{2\delta}}.
\]

Note that \( B(g_0, \varepsilon) \) (which is equal to the image of \( B(g_0, \varepsilon) \) in \( G/(MAN^-) \)) and \( B(g_0, \varepsilon) \) (which is equal to the image of \( B(g_0, \varepsilon) \) in \( G/(MAN^+) \)). Hence for all \( g \in B(g_0, \varepsilon) \), we have

\[
\nu_{g_0(o)}(u^+) = g_0 N_\varepsilon u^+_o \quad \text{and} \quad \nu_{g_0(o)}(u^-) = g_0 N_\varepsilon u^-_o
\]

where \( \varepsilon > 0 \). Moreover, for all \( g \in B(g_0, \varepsilon) \), \( \{ t \in \mathbb{R} : g a_t \in B(g_0, \varepsilon) \} \) has length precisely \( 2\varepsilon \) (see Lemma 2.1). Therefore the claim follows, since the BMS measure on \( G \) is the \( M \)-invariant extension of the BMS measure of \( G/M \).

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

\[
V_T(g_0, \varepsilon, \Omega) := B(g_0, \varepsilon)A_T^+ \Omega B(g_0, \varepsilon)^{-1}.
\]

(4.13)

We set

\[
V_T(\varepsilon, \Omega) := V_T(e, \varepsilon, \Omega)
\]

and note that

\[
V_T(g_0, \varepsilon, \Omega) := g_0 V_T(\varepsilon, \Omega) g_0^{-1}.
\]

Lemma 4.8. For all large \( T \gg 1 \) and small \( 0 < \varepsilon < 1 \), we have

\[
S_T(N_\varepsilon^+, (N_\varepsilon^-)^{-1}, \Omega) \subset V_T(\varepsilon, \Omega) \subset S_{T+\varepsilon}(N_\varepsilon^+, (N_\varepsilon^-)^{-1}, (N_{\varepsilon-e^+})^+\varepsilon^-)^{-1}, \Omega_\varepsilon^+)
\]

where \( \Omega_\varepsilon^+ = \cup_{m_1, m_2} m_1 \Omega m_2 \).

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

\[
g = g_+ g_0 g_- \in N^+(AM)N^-.
\]

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

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

Hence

\[
S_T(N_\varepsilon^+, (N_\varepsilon^-)^{-1}, \Omega) \subset V_T(\varepsilon, \Omega).
\]

(4.14)
On the other hand, if $g_1 \in \mathfrak{B}(\varepsilon), g_2 \in \mathfrak{B}(\varepsilon)^{-1}, a \in A_T^+, m \in M$, then

$$g_1 amg_2 \in (g_1)_+(N^+_{\varepsilon^{-T}}) amA_\varepsilon M_\varepsilon(N^-_{\varepsilon^{-T}})^{-1}(g_2)_-. $$

Therefore

$$V_T(\varepsilon, \Omega) \subset S_{T+\varepsilon}(N^+_{\varepsilon+\varepsilon^{-T}}, (N^-_{\varepsilon-\varepsilon^{-T}})^{-1}, \Omega^+). \tag{4.15}$$

This proves the claim.

**Theorem 4.9.** Let $\varepsilon > 0$ be smaller than the injectivity radius of $g_0$. We have

$$\# \Gamma \cap V_T(g_0, \varepsilon, \Omega) = (1 + O(\varepsilon)) \frac{e^{\delta T}}{\delta \cdot 2\varepsilon \cdot |m_{\text{BMS}}|} \cdot \left( m_{\text{BMS}}(\mathfrak{B}(g_0, \varepsilon) \otimes \Omega) + o(1) \right)$$

where the implied constants are independent of $\varepsilon$.

Moreover if $G$ and $\Gamma$ 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 4.8, we have

$$g_0 S_{T}(N^+_{\varepsilon}, (N^-_{\varepsilon})^{-1}, \Omega) g_0^{-1} \subset V_T(g_0, \varepsilon, \Omega) \subset g_0 S_{T}(N^+_{\varepsilon+\varepsilon^{-T}}, (N^-_{\varepsilon-\varepsilon^{-T}})^{-1}, \Omega^+ g_0^{-1}. $$

By Theorem 4.3 and Lemma 4.7, we have

$$\# \Gamma \cap V_T(g_0, \varepsilon, \Omega)
\begin{align*}
&= (1 + o(1)) \cdot \nu_{g_0/(o))((g_0 N^+_{\varepsilon})v_0^+)) \nu_{g_0/(o))((g_0 N^-_{\varepsilon})v_0^-)} \cdot \text{Vol}(\Omega) \delta^{-1} e^{\delta T} \\
&= (1 + O(\varepsilon)) (2\varepsilon)^{-1} \delta^{-1} e^{\delta T} \cdot (m_{\text{BMS}}(\mathfrak{B}(g_0, \varepsilon) \otimes \Omega) + o(1))
\end{align*}
$$

implying the first claim. The second claim follows from Theorem 4.6, and Remark 4.5.

**5 Asymptotic distribution of closed geodesics with holonomies**

We keep the notations $G, \Gamma, X, K, o, v_o$ etc. from Sect. 3. In particular, $\Gamma$ is Zariski dense and $|m_{\text{BMS}}| < \infty$, $X = \Gamma \setminus G/K$, and $T^1(X) = \Gamma \setminus G/M$. 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 Sect. 4. The main ingredient is the comparison Lemma 5.5, which we obtain using the effective closing Lemma 3.2.

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

$$\Gamma \setminus \Gamma gAM/M = \Gamma \setminus \Gamma gA(v_o)$$

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

For each closed geodesic $C$, we denote by $L_C$ 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 $\text{Lie}(S)/W$ where $S$ is a maximal torus of $M$ and $W$ is the Weyl group relative to $S$. For $T > 0$, define

$$G_T(T) := \{ C : C \text{ is a closed geodesic in } T^1(X), \, \ell(C) \leq T \}.$$

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

$$\mu_T(f \otimes \xi) = \sum_{C \in G_T(T)} L_C(f) \xi(h_C).$$

We also define a measure $\eta_T$ by

$$\eta_T(f \otimes \xi) = \sum_{C \in G_T(T)} D_C(f) \xi(h_C),$$

where $D_C(f) = \ell(C)^{-1} L_C(f)$. If $B$ is a subset of $\Gamma \backslash G/M$ and $\Omega$ 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 $\Gamma$ be geometrically finite and Zariski dense. For any bounded $f \in C(\Gamma \backslash G/M)$ and $\xi \in \text{Cl}(M)$, we have, as $T \to \infty$,

$$\mu_T(f \otimes \xi) \sim \frac{e^{\delta T}}{\delta |m^{BMS}|} \cdot m^{BMS}(f \otimes \xi); \quad (5.1)$$

and

$$\eta_T(f \otimes \xi) \sim \frac{e^{\delta T}}{\delta \cdot |m^{BMS}| \cdot T} \cdot m^{BMS}(f \otimes \xi). \quad (5.2)$$

Moreover if $G$ and $\Gamma$ are as Theorem 1.2, then both (5.1) and (5.2) 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 $\xi$.

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 $\Omega$ of $M^c$ and $g_0 \in G$. Recall the flow box $\mathcal{B}(g_0, \varepsilon) = g_0(N_+N^- \cap N_-^+N^+AM)M_\varepsilon A_\varepsilon$ and the notation $\tilde{\mathcal{B}}(g_0, \varepsilon) = \tilde{\pi}(\mathcal{B}(g_0, \varepsilon))$ from (4.11) and (4.12). We will first investigate the measure $\mu_T$ restricted to the set
The main idea is to relate the measure \( \mu_T(\mathfrak{B}(g_0, \varepsilon) \otimes \Omega) \) 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 5.6. 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 generated by a primitive hyperbolic element of \( \Gamma \), and \( C \) can be identified with \( \Gamma_{\tilde{C}} \backslash \tilde{C} \). Set

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

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 5.2.**

1. For any closed geodesic \( C \subset T^1(X) \), we have
   \[
   \mathcal{L}_C(\mathfrak{B}(g_0, \varepsilon)) = 2\varepsilon \cdot \# I(C).
   \]

2. For any \( T > 0 \), we have
   \[
   \mu_T(\mathfrak{B}(g_0, \varepsilon) \otimes \Omega) = 2\varepsilon \cdot \sum_{C \in \mathcal{G}_T(T)} \# I(C) \cdot 1_\Omega(h_C)
   \]
   where \( h_C \) is the holonomy class about \( C \).

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

\[
\mathcal{L}_C(\mathfrak{B}(g_0, \varepsilon)) = \int_{[ga_tv_o] \in \Gamma_{\Gamma_{\tilde{C}}} \backslash \tilde{C}} \sum_{\sigma \in \Gamma} 1_{\mathfrak{W}(g_0, \varepsilon)}(\sigma ga_tv_o)dt
\]

\[
= \sum_{[\sigma] \in \Gamma / \Gamma_{\tilde{C}}} \int_{[ga_tv_o] \in \tilde{C}} 1_{\mathfrak{W}(g_0, \varepsilon)}(\sigma ga_tv_o)dt.
\]

By Lemma 2.1, we have

\[
\int_{\tilde{C}} 1_{\mathfrak{W}(g_0, \varepsilon)}(\sigma ga_tv_o)dt = \begin{cases} 2\varepsilon, & \text{if } \sigma \tilde{C} \cap \mathfrak{W}(g_0, \varepsilon)v_o \neq 0 \\ 0, & \text{otherwise}. \end{cases}
\]

Therefore the claim follows. \( \square \)

Set

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

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

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

We denote by \( \Gamma_h \) the set of hyperbolic elements and by \( \Gamma_{ph} \) the set of primitive hyperbolic elements of \( \Gamma \).
PROPOSITION 5.3. For all large $T \gg 1$, we have
\[
\mu_{T}(\mathfrak{B}(g_0, \varepsilon) \otimes \mathfrak{W}_T(g_0, \varepsilon, \Omega)) = 2\varepsilon \cdot \#\Gamma_{ph} \cap \mathfrak{W}_T(g_0, \varepsilon, \Omega).
\]

Proof. We use Lemma 5.2 (2):
\[
\mu_{T}(\mathfrak{B}(g_0, \varepsilon) \otimes \mathfrak{W}_T(g_0, \varepsilon, \Omega)) = 2\varepsilon \cdot \sum_{C \in \mathcal{G}_T(T)} \#I(C) \cdot 1_{\Omega}(h_C).
\]
with $I(C) = \{\sigma(C') : \sigma C \cap \mathfrak{B}(g_0, \varepsilon)v_o \neq \emptyset\}$.

Upper bound: Let $C \in \mathcal{G}_T(T)$ 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 $[\sigma] \in I(C)$,
\[
\sigma_{\gamma} := \sigma \gamma \sigma^{-1} \in \mathfrak{W}_T(g_0, \varepsilon, \Omega);
\]
note that $\sigma_{\gamma}$ is well-defined independent of the choice of a representative $\sigma$ 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 $g_1v_o \in \tilde{C}$, and $\gamma = g_1\alpha_{\gamma}m_{\gamma}g_1^{-1}$ where $d(\alpha_{\gamma}, e) = \ell(C) \leq T$ 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
\[
g_2 a_{\gamma} m^{-1}_{\gamma} m_{\gamma} m = \sigma \gamma \sigma^{-1} g_2
\]
and
\[
\sigma_{\gamma} = g_2 a_{\gamma} m^{-1}_{\gamma} m_{\gamma} g_2^{-1} \in \mathfrak{W}_T(g_0, \varepsilon, \Omega)
\]
proving (5.6).

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

Lower bound: We write
\[
\#\Gamma_{ph} \cap \mathfrak{W}_T(g_0, \varepsilon, \Omega) = \sum \#[\gamma] \cap \mathfrak{W}_T(g_0, \varepsilon, \Omega)
\]
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 $\Gamma$. Fix a primitive hyperbolic element $\gamma \in \mathfrak{W}_T(g_0, \varepsilon, \Omega)$. So there exists $g \in \mathfrak{B}(g_0, \varepsilon)$ such that $\gamma = ga_{\gamma} m_{\gamma} g^{-1}$ with $a_{\gamma} \in A_T^+$ and $[m_{\gamma}] \in \Omega$. Let $C = \Gamma \setminus \Gamma gaAv_o$ and $\tilde{C} = gAv_o$. Then the length of $C$ is at most $T$.

For each element $\sigma' := \sigma \gamma \sigma^{-1} \in [\gamma] \cap \mathfrak{W}_T(g_0, \varepsilon, \Omega)$, 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 $\sigma^{-1}g_2 Av_o$ is the oriented axis for $\gamma$, $\sigma^{-1}g_2 Av_o = \tilde{C}$ by Corollary 2.3. Therefore $g_2v_o \in \sigma(\tilde{C}) \cap \mathfrak{B}(g_0, \varepsilon)v_o$, 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 (5.4).  
$\Box$
Indeed the proof of Proposition 5.3 gives that if $C$ is a closed geodesic and $[\gamma_C]$ is the conjugacy class of primitive hyperbolic elements which corresponds to $C$, then
\[
L_C(\mathfrak{B}(g_0, \varepsilon))1_\Omega(h_C) = 2\varepsilon \cdot #I(C) \cdot 1_\Omega(h_C) = 2\varepsilon \cdot #[\gamma_C] \cap \mathcal{W}(g_0, \varepsilon, \Omega).
\]  

**Lemma 5.4.** For $T > 1$, we have
\[
#\Gamma \cap (\mathcal{W}_T(g_0, \varepsilon, \Omega) - \mathcal{W}_{T/2}(g_0, \varepsilon, \Omega)) \leq #\Gamma_{ph} \cap \mathcal{W}_T(g_0, \varepsilon, \Omega).
\]

**Proof.** Note that, since $\mathcal{W}_T(g_0, \varepsilon, \Omega)$ consists of hyperbolic elements,
\[
#\Gamma_{ph} \cap \mathcal{W}_T(g_0, \varepsilon, \Omega) = #\Gamma \cap \mathcal{W}_T(g_0, \varepsilon, \Omega) - \#(\bigcup_{k \geq 2} \Gamma_k^{ph}) \cap \mathcal{W}_T(g_0, \varepsilon, \Omega)
\]
where $\Gamma_k^{ph} = \{\sigma^k : \sigma \in \Gamma_{ph}\}$.

Suppose $\gamma \in \mathcal{W}_T(g_0, \varepsilon, \Omega) \cap \Gamma_k^{ph}$ 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 $\gamma$ is hyperbolic, so is $\sigma$, and $\gamma$ and $\sigma$ 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
\[
#\mathcal{W}_T(g_0, \varepsilon, \Omega) \cap (\bigcup_{k \geq 2} \Gamma_k^{ph}) \leq #\Gamma \cap \mathcal{W}_{T/2}(g_0, \varepsilon, \Omega).
\]
Therefore the lower bound now follows. \hfill \Box

By the ergodicity of the geodesic flow with respect to the BMS measure on $\Gamma \backslash G$ [Win14], for any $g_0 \in \text{supp}(m_{\text{BMS}})$, 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 $\text{supp}(m_{\text{BMS}})$ 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 3.2 and the constant in (2.4).
Lemma 5.5 (Comparison lemma). For all \( T \gg 1 \), we have

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

where \( \Omega^- = \bigcap_{m_i \in M, c_\varepsilon} m_1 \Omega m_2 \).

Proof. The upper bound is immediate from the definition of the sets and Proposition 5.3. The effective closing Lemma 3.1 implies that for all large \( T \gg 1 \),

\[
\mathcal{V}_T(g_0, \varepsilon(1 - ce^{-T/2}), \Omega^-) - \mathcal{V}_{T/2}(g_0, \varepsilon, \Omega) \subset \mathcal{W}_T(g_0, \varepsilon, \Omega).
\]

Together with Proposition 5.3 and Lemma 5.4, this implies the lower bound.

Theorem 5.6. We have

\[
\mu_T(\tilde{\mathcal{B}}(g_0, \varepsilon) \otimes \Omega) = (1 + O(\varepsilon)) \frac{e^{\delta T}}{\delta \cdot |m_{\text{BMS}}|} \cdot (m_{\text{BMS}}(\tilde{\mathcal{B}}(g_0, \varepsilon) \otimes \Omega) + o(1))
\]

where the implied constants are independent of \( g_0 \) and \( \varepsilon \).

Moreover if \( G \) and \( \Gamma \) 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 5.5 and Theorem 4.9. \( \square \)

We note that we do not require \( \Gamma \) to be geometrically finite in the following theorem.

Theorem 5.7. Let \( \Gamma \) be Zariski dense with \( |m_{\text{BMS}}| < \infty \). For any \( f \in C_c(\Gamma \backslash G/M) \) and \( \xi \in \text{Cl}(M) \), we have, as \( T \to \infty \),

\[
\mu_T(f \otimes \xi) \sim \frac{e^{\delta T} \cdot m_{\text{BMS}}(f \otimes \xi)}{\delta \cdot |m_{\text{BMS}}|}.
\]

Moreover if \( G \) and \( \Gamma \) are as in Theorem 1.2, then (5.10) 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 \( \xi \).

Proof. We normalize \( |m_{\text{BMS}}| = 1 \). Using a partition of unity argument, we can assume without loss of generality that \( f \) is supported on \( \tilde{\mathcal{B}}(g_0, \varepsilon) \) 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 \( \tilde{\mathcal{B}}(h, \rho) \)'s with \( h \in \tilde{\mathcal{B}}(g_0, \varepsilon) \). Now applying Proposition 5.6 to each \( 1_{\tilde{\mathcal{B}}(h, \rho)} \otimes 1_\Omega \), we deduce that

\[
(1 - c\rho)m_{\text{BMS}}(f \otimes \xi) \leq \liminf_T \delta e^{-\delta T} \mu_T(f \otimes \xi) \leq \limsup_T \delta e^{-\delta T} \mu_T(f \otimes \xi) \leq (1 + c\rho)m_{\text{BMS}}(f \otimes \xi).
\]
Since ρ > 0 is arbitrary, this implies the claim when ξ is the characteristic function of Ω whose boundary has a measure zero. Via the identification $M^c = \text{Lie}(S)/W$ 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 (5.10). When the effective version of Theorem 5.6 holds, we also obtain an error term in this argument. ⊓⊔

**Contribution of the cusp and equidistribution for bounded functions.** In order to extend Theorem 5.7 to bounded continuous functions, which are not necessarily compactly supported, we now assume that $\Gamma$ is geometrically finite and use the following theorem of Roblin [Rob03] (Theorem 5.9).

We denote by $C(\Gamma)$ the convex core of $\Gamma$. Let $\varepsilon_0 > 0$ be the Margulis constant for $\Gamma$. Then $\{x \in C(\Gamma) : \text{injectivity radius at } x \geq \varepsilon_0\}$ is called the thick part and its complement is called the thin part. We will denote them $C(\Gamma)_{\text{thick}}$ and $C(\Gamma)_{\text{thin}}$ respectively. When $\Gamma$ is a geometrically finite group, the thin part part consists of finitely many disjoint cuspidal regions (called horoballs), say, $\mathcal{H}_1, \ldots, \mathcal{H}_k$ based at parabolic fixed points $p_1, \ldots, p_k$ respectively. We denote by $\Gamma_{p_i}$ the stabilizer of $p_i$ in $\Gamma$. Also, fixing $o$ in the thick part of $C(\Gamma)$, 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 5.8 ([DOP00]).** If $\Gamma$ is geometrically finite, then for each parabolic fixed point $p_i \in \Lambda(\Gamma)$, we have

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

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 $\text{cusp}(r) = \cup_i \mathcal{H}_i(r)$.

**Theorem 5.9** (Roblin [Rob03]). There exist absolute constants $c_0, c_1 > 0$ such that for any $T \gg 1$,

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

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

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

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

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

**Proof of Theorem 5.1.** We may assume $|m^{\text{BMS}}| = 1$. By Proposition 5.8,

$$\sum_{\sigma \in \Gamma_{p_i}, d(q_i, \sigma q_i) > s} d(q_i, \sigma q_i) e^{-\delta d(q_i, \sigma q_i)} \to 0.$$
as \( s \to \infty \). Therefore by Theorem 5.9,
\[
e^{-\delta T} \cdot \mu_T(\text{cusp}(r)K) = o_r(1).
\]
If we denote by \( \Phi_r \) a continuous approximation of the unit neighborhood of \( C(\Gamma) - (\cup H_i(r)) \) (that is, \( \Phi_r = 1 \) on the neighborhood and 0 outside a slightly bigger neighborhood) then Theorem 5.7 implies that
\[
e^{-\delta T} \delta \mu_T(f \cdot \Phi_r \otimes \xi) = m^{\text{BMS}}(f \cdot \Phi_r \otimes \xi) + o_T(1).
\]
Hence
\[
|e^{-\delta T} \delta \mu_T(f \otimes \xi) - m^{\text{BMS}}(f \otimes \xi)| = o_T(1) + o_r(1) + m^{\text{BMS}}(\text{cusp}(r)K)
\]
since the support of \( m^{\text{BMS}} \) is contained in \( C(\Gamma) \). By taking \( r \to \infty \), we finish the proof of the first claim (5.1). In view of (5.11) and Theorem 5.7, the claim on the error term follows as well.

We now deduce (5.2) from (5.1); this is done in [Rob03, 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
\[
\delta T e^{-\delta T} \eta_T(f \otimes \xi) \geq \delta e^{-\delta T} \mu_T(f \otimes \xi).
\] (5.12)
Therefore (5.1) implies that
\[
\liminf_T \delta T e^{-\delta T} \eta_T(f \otimes \xi) \geq m^{\text{BMS}}(f \otimes \xi).
\]
We now bound \( \eta(f \otimes \xi) \) from above. Let \( \varepsilon > 0 \) be a small fixed number. We have
\[
\delta T e^{-\delta T} \eta_T(f \otimes \xi) = \delta T e^{-\delta T} \left( \sum_{G_T((1-\varepsilon)T)} D_C(f)\xi(h_C) + \sum_{G_T - G_T((1-\varepsilon)T)} D_C(f)\xi(h_C) \right)
\leq \delta T e^{-\delta T} \left( \sum_{G_T((1-\varepsilon)T)} \ell(C)D_C(f)\xi(h_C) + \sum_{G_T - G_T((1-\varepsilon)T)} \frac{\ell(C)}{(1-\varepsilon)T}D_C(f)\xi(h_C) \right)
\leq T e^{-\delta \varepsilon T} \left( \delta e^{-\delta ((1-\varepsilon)T)}\mu_{(1-\varepsilon)T}(f \otimes \xi) + \frac{\delta e^{-\delta T}}{1-\varepsilon} \left( \mu_T(f \otimes \xi) - \mu_{(1-\varepsilon)T}(f \otimes \xi) \right) \right).
\]
Therefore again by Theorem 5.1,
\[
\limsup_T \delta T e^{-\delta T} \eta_T(f \otimes \xi) \leq m^{\text{BMS}}(f \otimes \xi) \left( \limsup_T T e^{-\delta \varepsilon T} + \frac{1}{1-\varepsilon} + \varepsilon^{-\delta T} \right)
\leq \frac{1}{1-\varepsilon} m^{\text{BMS}}(f \otimes \xi).
\]
Since \( \varepsilon > 0 \) is arbitrary, this proves the claim. When (5.1) is effective, the above argument also gives an effective statement (instead of using \( \limsup \) and \( \liminf \), we compare the error term and \( \varepsilon \) and take \( \varepsilon \) appropriately).
Acknowledgments

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.

References

[Bab02] M. Babillot. On the mixing property for hyperbolic systems. Israel J. Math., 129 (2002), 61–76
[BGS11] J. Bourgain, A. Gamburd and P. Sarnak. Generalization of Selberg’s 3/16 theorem and Affine sieve. Acta Math., 207 (2011), 255–290
[BKS10] J. Bourgain, A. Kontorovich and P. Sarnak. Sector estimates for Hyperbolic isometries. GAFA 20 (2010), 1175–1200
[Bow95] B. H. Bowditch. Geometrical finiteness with variable negative curvature. Duke Math. J., 77 (1995), 229–274
[CI99] K. Corlette and A. Iozzi. Limit sets of discrete groups of isometries of exotic hyperbolic spaces. Transaction of the AMS, 351 (1999), 1507–1530
[DOP00] F. Dal’bo, J.-P. Otal and M. Peigné. Sérries de Poincaré des groupes géométriquement finis. Israel J. Math., 118 (2000), 109–124
[FS90] L. Flaminio and R. Spatzier. Geometrically finite groups, Patterson–Sullivan measures and Ratner’s theorem. Inventiones, 99 (1990), 601–626
[GW80] R. Gangolli and G. Warner. Zeta functions of Selberg’s type for some non-compact quotients of symmetric spaces of rank one. Nagoya Math. J., 78 (1980), 1–44
[GO07] A. Gorodnik and H. Oh. Orbits of discrete subgroups on a symmetric space and the Furtstenberg boundary. Duke Math., 139 (2007), 483–525
[GOS10] A. Gorodnik, H. Oh and N. Shah. Strong wavefront lemma and counting lattice points in sectors. Israel J. Math., 176 (2010), 419–444
[Mar04] G. Margulis. On some aspects of the theory of Anosov systems. In: Springer Monographs in Mathematics. Springer-Verlag, Berlin (2004).
[MO14] A. Mohammadi and H. Oh. Matrix coefficients, counting and primes for orbits of geometrically finite groups. J. Eur. Math. Soc., arXiv:1208.4139 (2014, to appear)
[Nau05] F. Naud. Expanding maps on cantor sets and analytic continuation of zeta functions. Annales scientifiques de l’École normale supérieure, 38 (2005), 116–153
[OS13] H. Oh and N. Shah. Equisdistribution and Counting for orbits of geometrically finite hyperbolic groups. Journal of AMS, 26 (2013), 511–562.
[PP86] W. Parry and M. Pollicott. The Chebotarov theorem for Galois coverings of Axiom A flows. ETDS, 6 (1986), 133–148
[Pat76] S. Patterson. The limit set of a Fuchsian group. Acta Mathematica, 136 (1976), 241–273
[PR03] G. Prasad and A. Rapinchuk. Existence of irreducible ℝ-regular elements in Zariski-dense subgroups. Mathematical Research Letters, 10 (2003), 21–32
[Rob03] T. Roblin. Ergodicité et équidistribution en courbure négative. Mém. Soc. Math. Fr. (N.S.), 95 (2003), vi+96
[Sar83] P. Sarnak. The arithmetic and geometry of some hyperbolic three manifolds. Acta Math., (3–4)151 (1983), 253–295
[SW99] P. Sarnak and M. Wakayama. Equidistribution of holonomy about closed geodesics. Duke Math. J., 100 (1999), 1–57

[Sul79] D. Sullivan. The density at infinity of a discrete group of hyperbolic motions. Inst. Hautes Études Sci. Publ. Math., 50 (1979), 171–202

[Sul84] D. Sullivan. Entropy, Hausdorff measures old and new, and limit sets of geometrically finite Kleinian groups. Acta Math., (3–4)153 (1984), 259–277

[Vin14] I. Vinogradov. Effective bisector estimate with applications to Apollonian circle packings. Int Math Res Notices., (12)2014 (2014), 3217–3262

[Win14] D. Winter. Mixing of frame flow for rank one locally symmetric spaces and measure classification. arXiv:1403.2425 (Preprint)

GREGORY MARGULIS AND HEE OH, Mathematics Department, Yale University, New Haven, CT 06511, USA  hee.oh@yale.edu

Gregory Margulis  gregorii.margulis@yale.edu

AMIR MOHAMMADI, Department of Mathematics, The University of Texas at Austin, Austin, TX 78750, USA  amir@math.utexas.edu

HEE OH, Korea Institute for Advanced Study, Seoul, Korea

Received: May 1, 2014
Accepted: June 30, 2014