

# QUANTUM ERGODICITY IN THE BENJAMINI–SCHRAMM LIMIT FOR LOCALLY SYMMETRIC SPACES

FARRELL BRUMLEY, SIMON MARSHALL, JASMIN MATZ, AND CARSTEN PETERSON

ABSTRACT. We prove that for almost all symmetric spaces  $X$  and for any sequence of compact locally symmetric spaces  $Y_n$  which is uniformly discrete, has a uniform spectral gap, and converges in the sense of Benjamini–Schramm to  $X$ , the joint eigenfunctions of all invariant differential operators on  $Y_n$  delocalize on average when their spectral parameters are taken to lie in a fixed spectral window.

## CONTENTS

1. Introduction	1
2. Sketch of proof and main technical theorems	4
3. Preliminaries	9
4. Spectral estimate	17
5. Geometric estimate: statement and reduction steps	25
6. Intersection volumes	33
7. Semi-dense root subsystems	38
8. Bounding the elementary spectral integral	43
9. Bounds for spherical functions	48
References	59

## 1. INTRODUCTION

The celebrated Quantum Ergodicity theorem of Snirelman [Sni74], Zelditch [Zel87], and Colin de Verdière [Ver85] states that, on a closed Riemannian manifold  $Y$  whose geodesic flow is ergodic, the  $L^2$ -mass of almost every Laplacian eigenfunction equidistributes in the limit of high frequency. This is an early example of the transference principle in semiclassical analysis, whereby high-energy Laplacian eigenfunctions inherit the dynamical properties of the underlying Hamiltonian system, in this case given by the geodesic flow.

More precisely, let  $\{\psi_j\}$  be an orthonormal basis of  $L^2(Y)$  consisting of eigenfunctions of the Laplacian. We write  $\Delta\psi_j = \mu_j^2\psi_j$ , where  $\mu_j \geq 0$  is the frequency, and  $N(M) = \#\{j : \mu_j \leq M\}$ .

---

*Date:* January 12, 2026.

F.B. is supported by the Institut Universitaire de France and ANR-FNS Grant ANR-24-CE93-0016. S.M. was supported by National Science Foundation grant DMS-1902173. J.M is supported by the Carlsberg Foundation grant no. CF21-0374. C.P. was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) Grant SFB-TRR 358/1 2023 - 491392403 and received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 101034255.

$\mu_j \leq M\}$ . Quantum Ergodicity affirms that, for any  $a \in C(Y)$ ,

$$\lim_{M \rightarrow \infty} \frac{1}{N(M)} \sum_{j: \mu_j \leq M} \left| \int_Y a |\psi_j|^2 d\text{vol} - \frac{1}{\text{vol}(Y)} \int_Y a d\text{vol} \right|^2 = 0.$$

In fact, the aforementioned authors proved a stronger version of the above statement in which the multiplication operator  $a$  is replaced with more general pseudodifferential operators.

An analog of Snirelman's theorem can be posed in the setting of locally symmetric spaces of non-compact type, which form a subclass of Riemannian manifolds of central interest in automorphic forms and harmonic analysis. To do so effectively, one must take into account the extra symmetries enjoyed by these spaces. Indeed, Snirelman's theorem does not always directly apply, since the geodesic flow on the cosphere bundle of a locally symmetric space  $Y$  is ergodic only in rank 1. Nevertheless, the *Weyl chamber flow* can serve as a suitable substitute: if  $r$  is the rank of  $Y$ , the Weyl chamber flow is an  $\mathbb{R}^r$ -action on the Weyl chamber bundle, whose orbits descend to immersed maximal flat subspaces of  $Y$ , the higher rank analog of geodesics. This action is ergodic with respect to the uniform measure, and its quantization yields a rank  $r$  algebra of commuting differential operators containing the Laplacian. One can then ask for the quantum ergodic properties of the joint eigenfunctions for this algebra as the multispectra goes to infinity, a point of view first advanced in [SV07].

In this paper we shall investigate quantum ergodicity for locally symmetric spaces  $Y$  under a different limiting procedure to that described above. Following the breakthrough results of [ALM15; LMS17], rather than fixing a single space  $Y$  and varying the multispectra, we shall prove a version of quantum ergodicity for a fixed spectral window and a sequence of locally symmetric spaces  $Y_n = \Gamma_n \backslash X$  which converge, in the sense of Benjamini–Schramm, to their common universal cover  $X$ . In so doing we will correct a significant error in the earlier work of the first and third authors [BM23], in which  $X$  was taken to have isometry group  $\text{SL}_n(\mathbb{R})$ , and greatly expand the scope of that paper. The precise result is stated in Theorem 1.1 below.

The proof of our main theorem requires the introduction of several new techniques and auxiliary estimates, both in harmonic analysis (bounds on spherical functions) and the geometry of higher rank symmetric spaces (bounds on intersection translates), which we believe to be of general interest. These results, and their role in the proof of Theorem 1.1, are discussed in detail in Section 2.

**1.1. Review of literature: the rank one case.** The framework for studying quantum ergodicity in the Benjamini–Schramm limit originates in the work of Anantharaman–Le Masson [ALM15] in the context of regular graphs. They considered sequences of  $(q+1)$ -regular graphs for which the adjacency operator (the discrete analogue of the Laplacian) has a uniform spectral gap and for which the number of short loops is small compared to the number of vertices; the latter condition is equivalent to the Benjamini–Schramm convergence of the sequence of graphs to the  $(q+1)$ -regular tree. They proved a form of quantum ergodicity for eigenfunctions of the adjacency operator with eigenvalue lying in some subinterval of  $[-2\sqrt{q}, 2\sqrt{q}]$ , which is the spectrum of the adjacency operator acting on the  $(q+1)$ -regular tree.

Analogous results were proven for hyperbolic surfaces by Le Masson–Sahlsten [LMS17], building on work of Brooks–Le Masson–Lindenstrauss [BLML16], the latter of which gave a new proof of the result of Anantharaman–Le Masson [ALM15]. Such results were extended to

all rank one locally symmetric spaces by Abert–Bergeron–Le Masson [ABM22]. Furthermore, Anantharaman–Sabri [AS19] proved analogous results in the context of arbitrary graphs.

**1.2. Our main result.** Let us recall some notation related to locally symmetric spaces, which will be necessary to state our main result.

Let  $G$  be a connected non-compact semisimple Lie group with finite center and  $K < G$  a maximal compact subgroup. Let  $\mathfrak{g}$  and  $\mathfrak{k}$  be their respective Lie algebras. Then  $K$  induces a Cartan involution with corresponding eigenspace decomposition  $\mathfrak{g} = \mathfrak{p} \oplus \mathfrak{k}$ . We endow  $X = G/K$  with the  $G$ -invariant metric coming from the Killing form, making  $X$  a Riemannian symmetric space. All groups isogenous to  $G$  yield the same symmetric space. Thus we may assume without loss of generality that  $G$  is a product of non-compact centerless connected simple real Lie groups.

Let  $\Gamma < G$  be an irreducible lattice, which we shall assume to be uniform throughout this paper. Then  $Y = \Gamma \backslash X$  is a compact locally symmetric space. We endow  $Y$  with the measure  $dx$  induced from the Riemannian volume form on  $X$ . Let  $D_G(X)$  be the ring of  $G$ -invariant differential operators on  $X$ . The action of  $D_G(X)$  descends to functions on  $Y$ , and the joint eigenfunctions of  $D_G(X)$  in  $L^2(Y)$  are called *Maass forms*. A Maass form determines an eigencharacter of  $D_G(X)$ , which can then be identified with a spectral parameter  $\lambda \in \mathfrak{a}_\mathbb{C}^*/W$  by the Harish-Chandra isomorphism. Here  $\mathfrak{a}$  is a maximal abelian subspace of  $\mathfrak{p}$ ,  $\mathfrak{a}_\mathbb{C}$  is its complexification, and  $W$  is the Weyl group of  $G$ . A Maass form is said to be *tempered* when its spectral parameter lies in  $\mathfrak{a}^*/W$ , the latter being identified with the  $D_G(X)$ -spectrum on  $L^2(X)$ . For example, when  $X = \mathbb{H}^n$  is hyperbolic  $n$ -space and  $Y = \Gamma \backslash \mathbb{H}^n$  is a compact hyperbolic manifold, a Maass form  $\psi$  is just a Laplacian eigenfunction  $\Delta\psi = \mu^2\psi$ , and if we write the eigenvalue as  $\mu^2 = \rho^2 + \lambda^2$ , with  $\rho = (n-1)/2$ , then the tempered case  $\lambda \in \mathbb{R}$  corresponds to  $\mu^2 \geq \rho^2$ .

We shall be interested in the  $L^2$ -mass distribution of Maass forms, when their spectral parameters are confined to a compact subset of  $\mathfrak{a}^*/W$  and the lattice  $\Gamma < G$  is allowed to vary. A sequence of locally symmetric spaces  $Y_n = \Gamma_n \backslash X$  as above is said to converge to  $X$ , in the sense of *Benjamini–Schramm*, if asymptotically almost all points in  $Y_n$  have arbitrarily large injectivity radius. Moreover, if the  $\Gamma_n$  are torsion free, we say that the sequence  $\Gamma_n < G$  is *uniformly discrete* if there is a universal lower bound on the global injectivity radii of  $Y_n$ .

With the above notation, our main theorem is the following.

**Theorem 1.1.** *Let  $G$  be a product of non-compact connected centerless simple real Lie groups,  $K$  be a maximal compact subgroup, and  $X = G/K$  be the associated symmetric space. Let  $\Gamma_n < G$  be a sequence of torsion free, cocompact, uniformly discrete, irreducible lattices. Suppose  $Y_n = \Gamma_n \backslash X$  Benjamini–Schramm converges to  $X$  as  $n \rightarrow \infty$ . Fix a simple factor  $G_1$  of  $G$  and suppose that*

- (1) *its root subsystem of reduced roots is of type  $A_n, B_n, C_n, D_n$  or  $E_7$ ,*
- (2) *the action of  $G_1$  on the orthocomplement  $L_0^2(\Gamma_n \backslash G)$  to the constant functions has a uniform spectral gap.*

*There exists a finite  $W$ -stable set of hyperplanes  $\{P_i\}$  in  $\mathfrak{a}^*$  such that, for any compact  $W$ -invariant subset  $\Omega \subset \mathfrak{a}^* \setminus \cup_i P_i$  with non-empty interior, the following holds. Let  $\{\psi_j^{(n)}\}$  be an orthonormal basis of  $L^2(Y_n)$  consisting of Maass forms with associated spectral parameters  $\lambda_j^{(n)}$ , and write  $N(\Omega, \Gamma_n) = \#\{j : \lambda_j^{(n)} \in \Omega\}$ . Then for any uniformly bounded sequence*

$a_n \in L^\infty(Y_n)$  we have

$$\lim_{n \rightarrow \infty} \frac{1}{N(\Omega, \Gamma_n)} \sum_{j: \lambda_j^{(n)} \in \Omega} \left| \int_{Y_n} a_n(x) |\psi_j^{(n)}(x)|^2 dx - \frac{1}{\text{vol}(Y_n)} \int_{Y_n} a_n(x) dx \right|^2 = 0.$$

In particular, the theorem holds for a non-compact simple Lie group  $G$  of rank at least 2 satisfying (1) and any uniformly discrete sequence of cocompact torsion free lattices  $\Gamma_n < G$  such that  $\text{vol}(Y_n) \rightarrow \infty$ . Indeed, in such a setting the uniform spectral gap assumption (2) is automatic by Property (T), and the Benjamini–Schramm convergence of  $Y_n = \Gamma_n \backslash X$  to  $X$  is automatic by [ABB+17, Theorem 1.5].

The exceptional hyperplanes  $\cup_i P_i$  in the statement of Theorem 1.1 can be taken to depend only on the root system of  $G$ . In particular, when the root subsystem of reduced roots for  $G$  is of type  $A_n, B_n$ , or  $C_n$ , Theorem 1.1 holds with  $\Omega$  a compact subset of the regular parameters  $\mathfrak{a}_{\text{reg}}^*$ . See Section 4, and, in particular, Propositions 4.3 and 4.7, for a more complete description of the  $\cup_i P_i$ .

The exclusion of types  $E_6, E_8, F_4$  and  $G_2$  from the theorem results from a certain combinatorial property of the root system not holding in these types (see Section 7). There are precisely eight simple groups/irreducible symmetric spaces excluded from our theorem. In the notation of [Hel01, p. 532] these are the groups

$$E\text{ I}, E\text{ II}, E\text{ III}, E\text{ VI}, E\text{ VIII}, E\text{ IX}, F\text{ I}, \text{ and } G.$$

Despite the failure of our techniques to treat such spaces, we believe Theorem 1.1 should remain valid for them.

**1.3. Review of literature: higher rank case.** The first investigation of quantum ergodicity in the Benjamini–Schramm limit in higher rank can be found in the work of the first and third authors [BM23] who focused on locally symmetric spaces associated to  $\text{SL}_n(\mathbb{R})$ . At a certain step in the proof, which we refer to as the *geometric bound*, they must bound the volume of a certain set in the symmetric space. A mistake in the geometric bound was found by the fourth author as part of his thesis work [Pet23a]. This mistake resulted in a gap in the proof of the main theorem of [BM23].

By modifying the techniques of [BM23], particularly those related to the geometric bound, the fourth author proved quantum ergodicity in the Benjamini–Schramm limit for the group  $\text{PGL}_3$  over a non-archimedean local field, in which case the role of the symmetric space is replaced by that of the Bruhat-Tits building, and the invariant differential operators are replaced by the spherical Hecke algebra [Pet23b]. In Section 2.2 we shall say more about the mistake in the geometric bound of [BM23], the method of “fixing” it for non-archimedean  $\text{PGL}_3$  in [Pet23b], and how these ideas play a role in the present paper.

**1.4. Acknowledgements.** We would like to thank Jean-Philippe Anker, particularly for his contributions to Section 6.1 and 6.2.

## 2. SKETCH OF PROOF AND MAIN TECHNICAL THEOREMS

The overall strategy of the proof of Theorem 1.1 follows the argument of Le Masson–Sahlsten [LMS17], which itself derives from [BLML16]. The subsequent works of [ABM22], [BM23], [Pet23a] all followed the same basic structure, with additional difficulties depending on the given context (higher dimensional hyperbolic, higher rank real and  $p$ -adic). The

main purpose of the present discussion is to recall the various steps of the argument which are common to the above works, point to the specific section in this paper where they are executed, and describe in detail the formidable geometric and analytic challenges special to our higher rank setting.

Our two main auxiliary results which surmount these difficulties are stated in Sections 2.2 and 2.3 below. Of these, the primary result is stated in Theorem 2.2; it provides sharp (up to logarithmic powers) upper bounds on intersection volumes of spherical shells in the globally symmetric space  $X$ , and addresses the issue raised in §1.3. We prove this geometric bound using harmonic analysis. For this, we prove new uniform bounds on spherical functions in Theorem 2.3, which provides the central ingredient to our analytic approach.

To simplify the exposition, we shall assume throughout Section 2 that  $G$  is simple.

**2.1. Spectral and geometric reduction steps.** The basic strategy to the proof of Theorem 1.1 is to introduce an averaging operator over expanding bi- $K$ -invariant sets  $S_{tH_0} \subset G$  to convert the distributional properties the  $L^2$ -mass of Maass forms into the mean ergodic properties of the  $S_{tH_0}$  (or their intersection translates).

**Reduction to mean-zero observables.** We begin by observing that it is enough to prove Theorem 1.1 for the class of observables which are orthogonal to constants. Indeed, if  $a_n$  is as in the theorem, then so is  $a_n - \frac{1}{\text{vol}(Y_n)} \int_{Y_n} a_n$ , so we may assume, without loss of generality, that  $\int_{Y_n} a_n = 0$ . In this case the proof of Theorem 1.1 reduces to showing

$$(1) \quad \frac{1}{N(\Omega, \Gamma_n)} \sum_{\lambda_j^{(n)} \in \Omega} |\langle a_n \psi_j^{(n)}, \psi_j^{(n)} \rangle|^2 \longrightarrow 0$$

as  $n \rightarrow \infty$ .

**Spectral properties of the propagator (Section 4).** In Section 4, we introduce the aforementioned expanding bi- $K$ -invariant sets  $S_{tH_0}$  in  $G$ . More precisely, given a non-zero  $H_0 \in \bar{\mathfrak{a}}_+$  and real parameters  $t, \epsilon_0 > 0$ , define the *spherical shell directed by  $H_0$*  by

$$S_{tH_0} := K \exp(B_{\mathfrak{a}}(tH_0, \epsilon_0)) K,$$

where  $B_{\mathfrak{a}}(H, r)$  denotes the Euclidean ball in  $\mathfrak{a}$  of radius  $r > 0$  centered at  $H \in \mathfrak{a}$ . We may associate with  $S_{tH_0}$  its averaging operator  $U_t$  on  $L^2(\Gamma \backslash G)$ , and we denote the corresponding self-adjoint time average by  $\mathbf{A}(\tau)$ . The latter is an integral operator with kernel given by  $\sum_{\gamma \in \Gamma} A(g, \gamma h)$ , where

$$(2) \quad A(\tau)(g, h) = \frac{1}{\tau} \int_{\tau}^{2\tau} e^{-2t\rho(H_0)} \int_{gS_{tH_0} \cap hS_{tH_0}} a(x) dx dt.$$

The primary goal of Section 4 is then to prove that one can essentially replace the matrix coefficient  $\langle a\psi_j, \psi_j \rangle$  in the spectral average (1) by the matrix coefficient  $\langle \mathbf{A}(\tau)\psi_j, \psi_j \rangle$ .

This indeed is the content of Theorem 4.1, which is valid in a wide degree of generality. Since Theorem 4.1 is essentially local in nature, no assumptions on  $\Gamma$  are necessary. Moreover, the non-zero directing element  $H_0$  can be taken to be arbitrary, provided one restricts to spectral parameters avoiding the exceptional hyperplanes  $\{P_i\}$  figuring in the statement of Theorem 1.1. To be precise, these exceptional hyperplanes arise in the proof of Proposition 4.6, and depend on  $H_0$  through the set  $\mathfrak{a}_{\text{bad}}^*$ , defined in (31).

**Reduction to the intersection volume bound (Section 5).** The outcome of Section 4 is to reduce Theorem 1.1 to a similar estimate on the Hilbert–Schmidt norm  $\|\mathbf{A}(\tau)\|_{\text{HS}}$ . We study the latter by realizing it as the  $L^2$ -norm of its kernel. We state in Theorem 5.1 the ultimate bound we prove on  $\|\mathbf{A}(\tau)\|_{\text{HS}}$ , which involves various geometric and spectral quantities controlled by the hypotheses in Theorem 1.1. We then show, using known Limit Multiplicity theorems recalled in §3.8, how Theorem 5.1 suffices to deduce Theorem 1.1.

As is evident from (2), the Hilbert–Schmidt norm of the operator  $\mathbf{A}(\tau)$  encodes the dynamical properties of the intersection of the spherical shells  $S_{tH_0}$  with their group translates. The most difficult step in the proof of Theorem 5.1 is to bound the volume of these intersections, a task we accomplish in Theorem 2.2 below. The bulk of Section 5 is then dedicated to the reduction of Theorem 5.1 to Theorem 2.2.

This reduction proceeds by bounding  $\|\mathbf{A}(\tau)\|_{\text{HS}}$  by means of a thick-thin decomposition of the locally symmetric space  $Y = \Gamma \backslash G/K$ . This resulting bound produces a main term, denoted as  $M(\tau)$  and defined in (44), and an error term, occurring in Lemma 5.2. The error term involves the volume of the thin part and the global injectivity radius of  $Y$ , as well as the support of the kernel function (2). The former are controlled by the hypotheses of Theorem 5.1 while the latter is determined in Corollary 5.5.

The analysis of the the main term  $M(\tau)$  is much more delicate. In Proposition 5.8 we use the Minkowski integral inequality and the Nevo ergodic theorem to set up an application of Theorem 2.2. Once this is inserted, we execute in §5.4 one last integral using a degenerate form of Brion’s formula to conclude the proof of Theorem 5.1.

**2.2. Bounds on intersection volumes (Sections 6–9).** It then remains to bound the volumes of intersection translates of expanding spherical shells. For this, we wish to find an  $H_0$  such that for all  $H \in \mathfrak{a}$  and  $t \gg 1$ , the volume of the intersection  $e^H S_{tH_0} \cap S_{tH_0}$  is “small”, in a suitable sense. We are able to do so for all non-compact simple groups satisfying condition (1) of Theorem 1.1, as we now describe.

Denote by  $\Phi$  the set of restricted roots for  $A$  in  $G$  and by  $\Phi_{\text{red}} \subset \Phi$  the subsystem of reduced roots (also known as indivisible roots). Let  $M$  be the centralizer of  $H_0$  in  $G$ ; then  $M$  is a standard Levi subgroup of  $G$ . Let  $\Phi_M \subset \Phi$  be the roots of  $A$  in  $M$ , and  $\Phi_{M,\text{red}} \subset \Phi_{\text{red}}$  the corresponding subsystem of reduced roots.

If  $\Sigma$  is an abstract reduced root system, in Section 7 we define a notion of a semistandard root subsystem  $\Sigma_0$  being *semi-dense* in  $\Sigma$ ; roughly speaking this means that for any semi-standard root subsystem  $\Psi \subset \Sigma$ ,  $\Sigma_0$  contains at least half the roots of  $\Psi$ . In case  $\Phi_{\text{red}}$  is of type  $A_n, B_n, C_n, D_n$  or  $E_7$ , we define certain  $H_0$  which we call *extremely singular* which have the property that  $\Phi_{M,\text{red}}$  is semi-dense in  $\Phi_{\text{red}}$ .

The following result summarizes the content of Proposition 7.5 and Lemmas 7.7–7.8.

### Proposition 2.1.

- (1) *An irreducible reduced root system  $\Sigma$  contains a semi-dense root subsystem if, and only if,  $\Sigma$  is of type  $A_n, B_n, C_n, D_n$  or  $E_7$ .*
- (2) *Let  $G$  be such that  $\Phi_{\text{red}}$  is of type  $A_n, B_n, C_n, D_n$  or  $E_7$  and let  $H_0$  be extremely singular. Then  $\Phi_{M,\text{red}}$  is a semi-dense root subsystem of  $\Phi_{\text{red}}$ .*

The volume of  $S_{tH_0}$  is asymptotically of size  $e^{2t\rho(H_0)}$ , where  $\rho$  is the half-sum of the positive roots. Moreover, as we shall see in Section 5.2, the spherical shell  $S_{tH_0}$  and its translate  $e^H S_{tH_0}$ , where  $H \in \bar{\mathfrak{a}}_+$ , do not intersect as soon as  $\rho(H) > t\rho(H_0 - w_0 H_0) = 2t\rho(H_0)$ . The

following bound interpolates between these two extremities, losing only a logarithmic factor of  $t$ .

**Theorem 2.2.** *Let  $G$  be a non-compact simple real Lie group. If  $H_0 \in \bar{\mathfrak{a}}_+$  is such that  $\Phi_{M,\text{red}}$  is semi-dense in  $\Phi_{\text{red}}$ , then there exists a non-negative integer  $k$  such that for all  $H \in \bar{\mathfrak{a}}_+$  and  $t \gg 1$ :*

$$(3) \quad \text{vol}(e^H S_{tH_0} \cap S_{tH_0}) \ll (\log t)^k e^{\rho(2tH_0 - H)}.$$

One may obtain such a bound in a straightforward way and for all choices of  $H_0$  at the cost of replacing the  $(\log t)^k$  factor by a polynomial in  $t$ . The fact that we have  $\log t$  factors rather than polynomial factors in  $t$  is critical in the proof, as is explained in Section 6.1.

In [BM23], the first and third authors needed to bound a similar quantity as the LHS of (3), but for a choice of  $H_0$  which was not extremely singular. They claimed a bound like the RHS but with the  $(\log t)^k$  factor replaced by a constant. However, the fourth author found a mistake in their argument and was able to show that, in the non-archimedean setting for  $\text{PGL}_3(F)$  and for the choice of  $H_0$  from [BM23] and specific choices of  $H$ , we have

$$\text{vol}(\varpi^H S_{tH_0} \cap S_{tH_0}) \gg tq^{\rho(2tH_0 - H)},$$

where  $F$  is a non-archimedean local field,  $\varpi^H$  is the image of a uniformizer  $\varpi$  under the dominant cocharacter  $H : F^\times \rightarrow A$  (with  $A$  a maximal  $F$ -split torus),  $\mathcal{O}$  is the ring of integers,  $q$  is the order of the residue field, and  $S_{tH_0} = \text{PGL}_3(\mathcal{O})\varpi^{tH_0}\text{PGL}_3(\mathcal{O})$ . However, if one takes an extremely singular  $H_0$ , one obtains  $\text{vol}(\varpi^H S_{tH_0} \cap S_{tH_0}) \ll q^{\rho(2tH_0 - H)}$ ; see [Pet23a]. This suggests that in the archimedean setting we should not expect (3) to hold for generic choices of  $H_0$ .

**2.3. Bounds for spherical functions (Section 9).** The proof of Theorem 2.2 relies crucially on a new bound for the spherical function. As this bound holds on a general semisimple group, and may be of independent interest, we shall state and prove it independently of Theorem 1.1 and Theorem 2.2.

Let  $G$  be any noncompact semisimple Lie group with finite center. We let  $\varphi_\lambda$  denote the spherical function on  $G$  with spectral parameter  $\lambda \in \mathfrak{a}_{\mathbb{C}}^*$ . See §3.4 for definitions. To state our bound, we first define

$$f_\alpha(H, \lambda) := \min(|\alpha(H)| + 1, |\langle \lambda, \alpha \rangle|^{-1} + 1),$$

where  $\alpha \in \Phi_{\text{red}}^+$ ,  $H \in \mathfrak{a}$ , and  $\lambda \in \mathfrak{a}_{\mathbb{C}}^*$ . We subsequently define

$$(4) \quad \Theta(H, \lambda) := \sum_{w \in W} \prod_{\alpha \in \Phi_{\text{red}}^+} f_\alpha(H, w\lambda).$$

Note that  $\Theta$  is  $W$ -invariant in both  $\lambda$  and  $H$ , and also satisfies  $\Theta(H, \lambda) \geq 1$ . Let  $\mathfrak{a}^*(\kappa)$  denote those elements in  $\mathfrak{a}_{\mathbb{C}}^*$  whose imaginary part is bounded in size by  $\kappa$ .

**Theorem 2.3.** *There are  $a, \kappa > 0$  such that the following holds. For all  $H \in \bar{\mathfrak{a}}_+$  and all  $\lambda \in \mathfrak{a}^*(\kappa)$ , we have*

$$\varphi_\lambda(e^H) \ll (1 + \|\lambda\|)^a \Theta(H, \lambda) \max_{w \in W} e^{-(\rho + w\Im\lambda)(H)}.$$

We shall prove Theorem 2.3 in Section 9. To aid the reader in understanding the statement of the theorem, in Section 9.1 we shall illustrate it in the case of  $\mathrm{SL}_2(\mathbb{C})$  and  $\mathrm{SL}(2, \mathbb{R})$ , as well as deriving some simpler consequences of it in general. Section 9.2 also contains a discussion of the relation between Theorem 2.3 and previous bounds for the spherical function.

When  $G$  is a complex group, Theorem 2.3 has an interesting ‘self-improving’ property which allows us to significantly strengthen it, subject to the condition that  $\lambda$  lies in  $\mathfrak{a}^*$ . In particular, we expect this bound to be sharp in all aspects subject to this tempered condition on  $\lambda$ . Moreover, this bound may be transferred to the Cartan motion group corresponding to  $G$ , using the well known link between the Cartan motion group and the semisimple group in the complex case, and we shall state it for both groups simultaneously.

The Cartan motion group associated to  $G$  is the semidirect product  $\mathfrak{p} \ltimes K$ , which acts on the Euclidean symmetric space  $\mathfrak{p}$ . For  $\lambda \in \mathfrak{a}_{\mathbb{C}}^*$ , we have the Euclidean spherical function  $\varphi_{\lambda}^E \in C^{\infty}(\mathfrak{p})$  (also known as the generalized Bessel function) defined by

$$(5) \quad \varphi_{\lambda}^E(Z) = \int_K e^{i\lambda(kZ)} dk, \quad Z \in \mathfrak{p},$$

where we have extended  $\lambda$  to an element of  $\mathfrak{p}^*$  by orthogonality. To recall the relationship between  $\varphi_{\lambda}^E$  and  $\varphi_{\lambda}$ , we let  $Q$  be the function defined on  $\mathfrak{a}$  by

$$Q(H) = \prod_{\alpha \in \Phi^+} \sinh(\alpha(H))/\alpha(H).$$

As  $Q$  is Weyl invariant, it extends to a function on  $\mathfrak{p}$ , which is equal to the Jacobian of the exponential map  $\exp : \mathfrak{p} \rightarrow G/K$ . We then have

$$\varphi_{\lambda}(e^Z) = CQ(Z)^{-1/2} \varphi_{\lambda}^E(Z)$$

for some constant  $C > 0$ , see for instance Ch. IV, Theorem 4.7 of [Hel00]. Our bounds for  $\varphi_{\lambda}$  and  $\varphi_{\lambda}^E$  are as follows.

**Theorem 2.4.** *Suppose that  $G$  is a complex group. Then for  $H \in \bar{\mathfrak{a}}_+$  and  $\lambda \in \mathfrak{a}^*$ , we have*

$$(6) \quad \varphi_{\lambda}^E(H) \ll \sum_{w \in W} \prod_{\alpha \in \Phi^+} (1 + |\alpha(H)\langle w\lambda, \alpha \rangle|)^{-1},$$

and

$$(7) \quad \varphi_{\lambda}(e^H) \ll e^{-\rho(H)} \prod_{\alpha \in \Phi^+} (|\alpha(H)| + 1) \sum_{w \in W} \prod_{\alpha \in \Phi^+} (1 + |\alpha(H)\langle w\lambda, \alpha \rangle|)^{-1}.$$

We note that

$$Q(H)^{-1/2} \asymp e^{-\rho(H)} \prod_{\alpha \in \Phi^+} (|\alpha(H)| + 1)$$

for  $H \in \bar{\mathfrak{a}}_+$ , so that the bounds (6) and (7) are equivalent. In Section 9.1 we explain why we expect Theorem 2.4 to be sharp based on an analysis of the oscillatory integral (5), and in Section 9.2 we discuss the relation between Theorem 2.4 and previous bounds for the spherical function on complex groups. We deduce Theorem 2.4 from Theorem 2.3 in Section 9.7.

### 3. PRELIMINARIES

In this section we introduce most of the notation that will be in force throughout the rest of the paper, and recall some of the definitions and foundational results that enter into the statement and proof of Theorem 1.1. More exactly, after setting up some standard notational conventions in §§3.1-3.4, we give standard bounds on the  $c$ -function and state the Harish-Chandra expansion in §3.5, we define *maximally singular* and *extremely singular* elements in §3.6, and we state the pertinent formulation of Benjamini–Schramm convergence in §3.7. Finally, in §§3.8-3.9 we recall the relevant limit multiplicity theorems and the notion of uniform spectral gap.

Throughout this section we shall let  $G$  denote a connected semisimple real Lie group with finite center and no compact factors. Additional hypotheses on  $G$  (such as those appearing in Theorem 1.1) will be assumed later in the paper as needed.

**3.1. Basic notation.** Fix a maximal compact subgroup  $K$  of  $G$ . Denote by  $\mathfrak{k}$  and  $\mathfrak{g}$  their respective Lie algebras. Then  $K$  induces a Cartan involution  $\Theta$  on  $G$ . The differential at the identity of  $\Theta$  defines an involution on  $\mathfrak{g}$ , whose  $-1$  and  $+1$  eigenspaces determine the Cartan decomposition  $\mathfrak{g} = \mathfrak{p} \oplus \mathfrak{k}$ .

Let  $\mathfrak{a} \subset \mathfrak{p}$  denote a maximal abelian subspace. Let  $A = \exp(\mathfrak{a})$  be the corresponding analytic subgroup of  $G$ . Let  $\mathfrak{a}^*$  denote the dual space of  $\mathfrak{a}$ . Let  $\Phi \subset \mathfrak{a}^*$  be the set of restricted roots. Let  $\Phi^+ \subset \Phi$  be a choice of positive roots. We let  $\Delta \subset \Phi^+$  denote the set of simple roots; they form a basis for  $\mathfrak{a}^*$ . Then the fundamental coweights  $\widehat{\Delta}^\vee = \{\varpi_\alpha^\vee : \alpha \in \Delta\} \subset \mathfrak{a}$  denote the dual basis to the simple roots, giving rise to decompositions  $\lambda = \sum_{\alpha \in \Delta} \lambda(\varpi_\alpha^\vee) \alpha$ ,  $H = \sum_{\alpha \in \Delta} \alpha(H) \varpi_\alpha^\vee$  of elements  $\lambda \in \mathfrak{a}^*$  and  $H \in \mathfrak{a}$ . Let

$$\mathfrak{a}_+ = \{H \in \mathfrak{a} : \alpha(H) > 0 \ \forall \alpha \in \Delta\} = \sum_{\alpha \in \Delta} \mathbb{R}_{>0} \varpi_\alpha^\vee$$

be the fundamental Weyl chamber in  $\mathfrak{a}$  and denote by  $\bar{\mathfrak{a}}_+ = \sum_{\alpha \in \Delta} \mathbb{R}_{\geq 0} \varpi_\alpha^\vee$  its closure.

We have a restricted root space decomposition

$$\mathfrak{g} = \mathfrak{m} \oplus \mathfrak{a} \oplus \sum_{\alpha \in \Phi} \mathfrak{g}_\alpha,$$

where  $\mathfrak{m} = Z_{\mathfrak{k}}(\mathfrak{a})$  is the centralizer of  $\mathfrak{a}$  in  $\mathfrak{k}$  and  $\mathfrak{g}_\alpha$  is the root space for  $\alpha$ . As usual, we let  $\rho = \frac{1}{2} \sum_{\alpha \in \Phi^+} m_\alpha \alpha$  be the half-sum of the positive roots, where we have put  $m_\alpha = \dim \mathfrak{g}_\alpha$ . Let  $r = \dim \mathfrak{a}$  be the rank of  $G$  and  $W = N_K(\mathfrak{a})/Z_K(\mathfrak{a})$  the Weyl group of  $G$ . Let  $\Phi_{\text{red}}$  denote the set of reduced roots of  $\Phi$ , i.e., roots  $\alpha$  such that  $\alpha/2$  is not also a root. Then  $\Phi_{\text{red}}$  is a reduced root system, and we refer to it as the reduced root system of  $G$ . Let  $\Phi_{\text{red}}^+ := \Phi_{\text{red}} \cap \Phi^+$ . Note that  $\Delta$  is also a base for  $\Phi_{\text{red}}$ .

We let  $\langle \cdot, \cdot \rangle$  denote the restriction of the Killing form to  $\mathfrak{a}$ . We equip  $\mathfrak{a}^*$  with the inner product  $\langle \alpha, \beta \rangle = \langle H_\alpha, H_\beta \rangle$ , where  $H_\alpha \in \mathfrak{a}$  is uniquely determined by  $\alpha(Z) = \langle H_\alpha, Z \rangle$ . For any non-zero  $\alpha \in \mathfrak{a}^*$ , we may then define the orthogonal reflection  $\lambda \mapsto \lambda - 2 \frac{\langle \lambda, \alpha \rangle}{\langle \alpha, \alpha \rangle} \alpha$  across the root hyperplane  $\alpha^\perp$ . Note that  $\lambda \mapsto 2 \frac{\langle \lambda, \alpha \rangle}{\langle \alpha, \alpha \rangle}$  is an element in  $\mathfrak{a}^{**}$ , and as such determines an element  $\alpha^\vee \in \mathfrak{a}$  such that  $2 \frac{\langle \lambda, \alpha \rangle}{\langle \alpha, \alpha \rangle} = \lambda(\alpha^\vee)$ . Then the orthogonal reflection can be written  $\lambda \mapsto \lambda - \lambda(\alpha^\vee) \alpha$  and hence  $\alpha^\perp = \{\lambda \in \mathfrak{a}^* : \lambda(\alpha^\vee) = 0\}$ .

Denote the set of (restricted) coroots and simple coroots by  $\Phi^\vee = \{\alpha^\vee : \alpha \in \Phi\} \subset \mathfrak{a}$  and  $\Delta^\vee = \{\alpha^\vee : \alpha \in \Delta\} \subset \mathfrak{a}$ , respectively. Then the fundamental weights  $\widehat{\Delta} = \{\varpi_\alpha : \alpha \in \Delta\} \subset$

$\mathfrak{a}^*$  form the dual basis to the simple coroots  $\Delta^\vee$ . We have corresponding decompositions  $\lambda = \sum_{\alpha \in \Delta} \lambda(\alpha^\vee) \varpi_\alpha$ ,  $H = \sum_{\alpha \in \Delta} \varpi_\alpha(H) \alpha^\vee$  of elements  $\lambda \in \mathfrak{a}^*$  and  $H \in \mathfrak{a}$ . Let

$$\mathfrak{a}_+^* = \{\lambda \in \mathfrak{a}^* : \lambda(\alpha^\vee) > 0 \ \forall \alpha \in \Delta\} = \sum_{\alpha \in \Delta} \mathbb{R}_{>0} \varpi_\alpha$$

be the fundamental Weyl chamber in  $\mathfrak{a}^*$  and denote by  $\bar{\mathfrak{a}}_+^* = \sum_{\alpha \in \Delta} \mathbb{R}_{\geq 0} \varpi_\alpha$  its closure. We write

$$(8) \quad \mathfrak{a}_{\text{sing}}^* = \bigcup_{\alpha \in \Phi^+} \{\lambda \in \mathfrak{a}^* : \lambda(\alpha^\vee) = 0\}$$

for the union of all root hyperplanes and put  $\mathfrak{a}_{\text{reg}}^* = \mathfrak{a}^* \setminus \mathfrak{a}_{\text{sing}}^*$ . Then  $\mathfrak{a}_{\text{reg}}^* = W \cdot \mathfrak{a}_+^*$ . Finally, we write  $\mathfrak{a}_{\mathbb{C}}^* := \mathfrak{a}^* \otimes \mathbb{C}$  for the complexification of  $\mathfrak{a}^*$ .

**3.2. Subsets of simple roots and associated structures.** Let  $\mathfrak{n} = \bigoplus_{\alpha \in \Phi^+} \mathfrak{g}_\alpha$  and write  $N = \exp(\mathfrak{n})$  for the corresponding analytic subgroup. Let  $P = Z_K(\mathfrak{a})AN$  be the standard minimal parabolic subgroup. The Bruhat decomposition tells us that

$$G = \bigsqcup_{w \in W} PwP,$$

where we choose any coset representative for  $w \in W = N_K(\mathfrak{a})/Z_K(\mathfrak{a})$ .

Any subgroup of  $G$  containing  $P$  is called a standard parabolic subgroup. These can be constructed in the following way. Let  $I \subseteq \Delta$ , and let  $W_I$  be the subgroup of  $W$  generated by the reflections associated to  $\alpha \in I$ ; we define  $W_\emptyset = 1$ . All standard parabolic subgroups are of the form

$$Q_I = \bigsqcup_{w \in W_I} PwP$$

for some  $I \subseteq \Delta$  [Hum75, §§29.2-3]. This recovers the Bruhat decomposition for  $G$  when  $I = \Delta$ .

Let  $\Phi_I$  be the root subsystem whose elements consist of roots in  $\Phi$  which are linear combinations of roots in  $I$ . Let  $\Phi_I^+ = \Phi_I \cap \Phi^+$ . Note that the roots of  $Q_I$  with respect to  $A$  are  $\Phi^+ \cup (-\Phi_I^+) = (\Phi^+ \setminus \Phi_I^+) \cup \Phi_I$  [Hum75, §30.1, Theorem (b)]. Since  $W_I$  preserves these roots, and it preserves  $\Phi_I$ , we immediately see that  $W_I$  preserves  $\Phi^+ \setminus \Phi_I^+$ . The following proposition relates  $\Phi_I^+$  with the action of  $W_I$  on  $\Phi^+$ .

**Proposition 3.1.** *Let  $I \subseteq \Delta$ . Then*

$$(9) \quad W_I = \{w \in W \mid w\Phi^+ \subseteq (\Phi^+ \setminus \Phi_I^+) \cup \Phi_I\}.$$

*In particular,  $w(\Phi^+ \setminus \Phi_I^+) = \Phi^+ \setminus \Phi_I^+$  if and only if  $w \in W_I$ . Furthermore,*

$$(10) \quad \begin{aligned} \Phi_I^+ &= \{\alpha \in \Phi^+ \mid \exists w \in W_I : w\alpha < 0\}, \\ \Phi^+ \setminus \Phi_I^+ &= \{\alpha \in \Phi^+ \mid w\alpha > 0 \ \forall w \in W_I\}. \end{aligned}$$

*Proof.* The direct inclusion in (9) holds, since  $W_I$  preserves the roots of  $Q_I$ . On the other hand, if  $w \in W$  is such that  $w\Phi^+ \subseteq \Phi^+ \cup (-\Phi_I^+)$ , then  $w$  maps  $P$  into  $Q_I$  and therefore belongs to  $Q_I$  by [Hum75, §29.3, Lemma D] so that  $w \in Q_I \cap W = W_I$ .

To see that  $W_I$  consists precisely of those  $w \in W$  for which  $w(\Phi^+ \setminus \Phi_I^+) = \Phi^+ \setminus \Phi_I^+$  we first observe that since  $W_I$  preserves  $\Phi_I$  and  $(\Phi^+ \setminus \Phi_I^+) \cup \Phi_I$ , it must also preserve  $\Phi^+ \setminus \Phi_I^+$ . Next, taking complements in (9) gives

$$W_I = \{w \in W \mid -(\Phi^+ \setminus \Phi_I^+) \subset w(-\Phi^+)\} = \{w \in W \mid w^{-1}(\Phi^+ \setminus \Phi_I^+) \subset \Phi^+\},$$

so that  $w(\Phi^+ \setminus \Phi_I^+) = \Phi^+ \setminus \Phi_I^+$  implies  $w \in W_I$  as required.

The two statements in (10) are clearly equivalent so that it suffices to prove the second. The direct inclusion follows from the fact that  $W_I$  preserves  $\Phi^+ \setminus \Phi_I^+$ . On the other hand, suppose  $\alpha \in \Phi^+$  is such that  $w\alpha > 0$  for all  $w \in W_I$ . It suffices to show that  $W_I\alpha \subset \Phi^+ \setminus \Phi_I^+$ , since this implies, by the preceding claim, that  $\alpha \in \Phi^+ \setminus \Phi_I^+$ . Now (9) already yields the inclusion  $W_I\alpha \subset (\Phi^+ \setminus \Phi_I^+) \cup \Phi_I$ . But  $W_I\alpha$  has no intersection with  $\Phi_I$ . Indeed, every element in  $W\alpha$  is positive while the  $W_I$ -orbit of every element in  $\Phi_I$  meets  $-\Phi^+$ , since  $\Phi_I$  is a root system with Weyl group  $W_I$ .  $\square$

Standard parabolic subgroups  $Q = Q_I$  admit a Levi decomposition

$$Q = L \ltimes U \quad (L = Q \cap \Theta(Q)),$$

where  $U$  is the unipotent radical of  $Q$ , whose Lie algebra is  $\bigoplus_{\alpha \in \Phi^+ \setminus \Phi_I^+} \mathfrak{g}_\alpha$ . (When  $I = \emptyset$ , we have  $Q = P$  and  $U = N$ .) The reductive subgroup  $L$ , whose Lie algebra decomposes as  $\mathfrak{a} \oplus \mathfrak{m} \oplus \bigoplus_{\alpha \in \Phi_I} \mathfrak{g}_\alpha$ , is called the Levi component of  $Q$ ; by *standard Levi subgroup* we shall mean any Levi component of a standard parabolic subgroup. The Weyl group of  $L$ , relative to the maximal compact subgroup  $K_L = K \cap L$ , and maximal split torus  $A \subset L$ , is equal to  $W_I$  [Hum75, §27.1, Theorem].

The Levi component  $L_I$  itself determines the standard parabolic subgroup  $Q_I$ . Moreover, we may identify  $L_I$  as the centralizer of any element in the open cone  $C_I = \sum_{\alpha \in \Delta \setminus I} \mathbb{R}_{>0} \varpi_\alpha^\vee$ , and all standard Levi subgroups are obtained in this way [Hum75, §30.2]. The nodes of the Dynkin diagram of  $\Phi$  are also naturally labeled by elements in  $\Delta$ . We thus have one-to-one correspondences between subsets  $I \subseteq \Delta$ , standard Levi subgroups  $L_I$ , standard parabolics  $Q_I$ , faces  $C_I$  of  $\bar{\mathfrak{a}}_+$ , and subdiagrams of the Dynkin diagram for  $\Phi$  obtained by deleting the nodes corresponding to  $\Delta \setminus I$ . In this last correspondence, the type of the resulting Dynkin diagram in fact tells us the type of the root system  $\Phi_I$ .

**3.3. Measures and Jacobian factors.** Let  $X = G/K$  be the Riemannian globally symmetric space associated to  $G$ . Let  $B$  denote the Killing form on  $\mathfrak{g}$ . Then  $B$  is positive definite on  $\mathfrak{p}$  and hence defines an inner product on  $\mathfrak{p}$ . Identifying  $\mathfrak{p}$  with the tangent space of  $X$  at the origin, we then transport this inner product to a  $G$ -invariant Riemannian metric on  $X$ , of non-positive curvature. The restriction of  $B$  to  $\mathfrak{a}$  defines a  $W$ -invariant inner product. Let  $dH$  denote the corresponding Lebesgue measure on  $\mathfrak{a}$ . Let  $dk$  denote the probability Haar measure on  $K$ . We normalize Haar measure  $dg$  on  $G$  in such a way that its quotient by  $dk$  agrees with the measure  $dx$  induced by the Riemannian volume form on  $X$ . If  $E$  is a subset of  $G$  or  $X = G/K$ , we write  $\text{vol}(E)$  for  $\int_E dg$  or  $\int_E dx$ , respectively.

The Cartan decomposition states that  $G = K \exp(\bar{\mathfrak{a}}_+) K$ . Correspondingly, there is a constant  $b_G > 0$  such that, for every  $f \in C_c(G)$ , we have

$$(11) \quad \int_G f(g) dg = b_G \int_K \int_{\mathfrak{a}_+} \int_K f(k_1 \exp(H) k_2) J(H) dk_1 dH dk_2,$$

where, for  $H \in \mathfrak{a}_+$ , the radial volume factor  $J \in C^\infty(\mathfrak{a}_+)$  is given by

$$(12) \quad J(H) = \prod_{\alpha \in \Phi^+} (\sinh \alpha(H))^{m_\alpha}.$$

See [GV88, §2.4, Prop. 2.4.11]. By extending to  $\mathfrak{a}_{\text{reg}}$  by  $W$ -invariance, and then by continuity from  $\mathfrak{a}_{\text{reg}}$  to  $\mathfrak{a}$ , we may also view  $J(H)$  as a continuous  $W$ -invariant function on  $\mathfrak{a}$ .

More generally, we let  $L = L_I$  be a standard Levi subgroup, where  $I$  is the corresponding subset of  $\Delta$ , as in the bijective correspondence of §3.2. Denote by  $W_L = W_I$  its Weyl group. Let  $\Phi_L = \Phi_I$ , and  $\Phi_L^+ = \Phi_L \cap \Phi^+$ . We let

$$(13) \quad \mathfrak{a}_{L,+} = \{H \in \mathfrak{a} : \alpha(H) > 0 \ \forall \alpha \in I\}$$

be the fundamental Weyl chamber for  $L$ . We again give  $K_L$  the probability Haar measure  $dk$ , and we denote by  $dl$  the Haar measure on  $L$ , normalized in the same way as for  $G$ . The Cartan decomposition for  $L$  states that  $L = K_L \exp(\bar{\mathfrak{a}}_{L,+}) K_L$ . Once again, there is a constant  $b_L > 0$ , depending only on  $L$ , such that

$$(14) \quad \int_L f(l) dl = b_L \int_{K_L} \int_{\mathfrak{a}_{L,+}} \int_{K_L} f(k_1 \exp(H) k_2) J_L(H) dk_1 dH dk_2,$$

where the radial volume factor with respect to  $L$  is

$$(15) \quad J_L(H) = \prod_{\alpha \in \Phi_L^+} (\sinh \alpha(H))^{m_\alpha}.$$

We define  $J^L = J/J_L$ .

**3.4. Spherical functions, spherical transform, spherical inversion.** The Iwasawa decomposition for  $G$  states that the multiplication map  $N \times A \times K \rightarrow G$  is a diffeomorphism. Let  $\mathcal{H} : G \rightarrow \mathfrak{a}$  be the Iwasawa projection, given by  $g = n \exp(\mathcal{H}(g)) k$ . For  $\lambda \in \mathfrak{a}_{\mathbb{C}}^*$  the map  $g \mapsto e^{i\lambda(\mathcal{H}(g))}$  descends to a plane wave on  $X$ . The average over left- $K$ -orbits

$$(16) \quad \varphi_\lambda(g) := \int_K e^{(i\lambda + \rho)(\mathcal{H}(kg))} dk$$

defines a bi- $K$ -invariant function called the spherical function with spectral parameter  $\lambda$ ; such functions can also be viewed as  $K$ -invariant functions on  $X$ . The spherical functions are in fact distinguished by the properties of equalling 1 at the identity, being bi- $K$ -invariant, and being joint eigenfunctions of all  $G$ -invariant differential operators on  $X$ , denoted  $D_G(X)$ . More precisely, the Harish-Chandra isomorphism gives an isomorphism

$$\gamma_{\text{HC}} : D_G(X) \xrightarrow{\sim} \text{Sym}(\mathfrak{a}_{\mathbb{C}})^W$$

onto the  $W$ -invariant elements of the symmetric algebra of  $\mathfrak{a}_{\mathbb{C}}$ , which we may in turn interpret as  $W$ -invariant polynomials on  $\mathfrak{a}_{\mathbb{C}}^*$ . We have that

$$D\varphi_\lambda = \gamma_{\text{HC}}(D)(\lambda)\varphi_\lambda$$

for all  $D \in D_G(X)$ . We have  $\varphi_\lambda = \varphi_{\lambda'}$  if and only if  $\lambda$  and  $\lambda'$  are in the same  $W$ -orbit. For this reason, we call  $\mathfrak{a}_{\mathbb{C}}^*/W$  the space of *spectral parameters*. The spectral parameters lying in  $\mathfrak{a}^*/W$  are called *tempered spectral parameters*. The tempered subspace  $\mathfrak{a}^*/W$  arises naturally in the decomposition of  $L^2(X)$ , as will be seen in the following subsection.

The group  $L$  has its own Iwasawa decomposition  $L = N_L K_L$  where  $N_L$  is the analytic subgroup corresponding to  $\bigoplus_{\alpha \in \Phi_L^+} \mathfrak{g}_\alpha$ . We can thus in turn define the spherical function  $\varphi_\lambda^L$  on  $L$  by the formula (16), replacing  $\rho$  with  $\rho_L = \frac{1}{2} \sum_{\alpha \in \Phi_L^+} m_\alpha \alpha$ . Then  $\varphi_\lambda^L = \varphi_{\lambda'}^L$  if and only if  $\lambda$  and  $\lambda'$  are in the same  $W_L$ -orbit.

Let  $C_c^\infty(G//K)$  denote the space of compactly supported smooth bi- $K$ -invariant functions. Given  $k \in C_c^\infty(G//K)$  let

$$(17) \quad \widehat{k}(\lambda) = \int_G k(g) \varphi_{-\lambda}(g) dg = b_G \int_{\mathfrak{a}^+} k(e^H) \varphi_{-\lambda}(e^H) J(H) dH$$

be the Harish-Chandra spherical transform. This transform can be inverted using the Harish-Chandra  $c$ -function  $c(\lambda)$ , which is a meromorphic function on  $\lambda \in \mathfrak{a}_{\mathbb{C}}^*$ , given by the product

$$(18) \quad c(\lambda) = C \prod_{\alpha \in \Phi_{\text{red}}^+} c_\alpha \left( \frac{i\langle \lambda, \alpha \rangle}{\langle \alpha, \alpha \rangle} \right), \quad c_\alpha(s) = \frac{2^{-s} \Gamma(s)}{\Gamma(\frac{1}{2}(\frac{1}{2}m_\alpha + 1 + s)) \Gamma(\frac{1}{2}(\frac{1}{2}m_\alpha + m_{2\alpha} + s))},$$

for a non-zero constant  $C$ . Then we have the following inversion formula:

$$k(g) = \int_{\mathfrak{a}_+^*} \widehat{k}(\lambda) \varphi_\lambda(g) |c(\lambda)|^{-2} d\lambda.$$

More generally, for a standard Levi subgroup  $L$ , we let  $c_L$  be its  $c$ -function, given by the product (18) but with  $\Phi_{\text{red}}^+$  replaced by  $\Phi_{\text{red},L}^+$ .

**3.5. Bounds on the  $c$ -function and the spherical function.** In this section we recall standard bounds on the  $c$ -function that will be used throughout the later sections and state the generalized Harish-Chandra expansion of the spherical function, a result which undergirds the proof of Theorem 2.3.

From its definition in (18) it is clear that each function  $c_\alpha(s)$  has a simple pole at  $s = 0$ . Therefore, for  $\lambda$  close to zero, we have

$$(19) \quad |c(\lambda)|^{-2} \ll \prod_{\alpha \in \Phi_{\text{red}}^+} |\langle \lambda, \alpha \rangle|^2.$$

On the other hand, if  $s$  is large with real part bounded by  $\kappa$ , we have that [DKV79]

$$(20) \quad |c_\alpha(s)|^{-2} \ll_\kappa |s|^{m_\alpha + m_{2\alpha}}.$$

Let  $L$  be a standard Levi subgroup of  $G$ . From the expression (18) it follows that  $c(\lambda)$  is holomorphic and non-vanishing on the regular parameters  $\mathfrak{a}_{\text{reg}}^*$ . Then the quotient

$$(21) \quad c^L(\lambda) = \frac{c(\lambda)}{c_L(\lambda)} = C^L \prod_{\alpha \in \Phi_{\text{red}}^+ \setminus \Phi_{\text{red},L}^+} c_\alpha \left( \frac{i\langle \lambda, \alpha \rangle}{\langle \alpha, \alpha \rangle} \right) \quad (C^L = C/C_L),$$

is holomorphic and non-vanishing on

$$(22) \quad \mathfrak{a}^* \setminus \bigcup_{\alpha \in \Phi^+ \setminus \Phi_L^+} \{ \lambda \in \mathfrak{a}^* : \lambda(\alpha^\vee) = 0 \},$$

an open set in  $\mathfrak{a}^*$  containing  $\mathfrak{a}_{\text{reg}}^*$ . The main term in the generalized Harish-Chandra expansion, stated below, will be seen to be

$$(23) \quad \theta_L(H, \lambda) := \sum_{w \in W_L \setminus W} c^L(w\lambda) \varphi_{w\lambda}^L(e^H),$$

and the error term is governed by the function  $\beta_L : \mathfrak{a} \rightarrow \mathbb{R}$  given by

$$(24) \quad \beta_L(H) := \min_{\alpha \in \Phi^+ \setminus \Phi_L^+} |\alpha(H)|.$$

Note that  $\theta_L(H, \lambda)$  is defined on  $\mathfrak{a}_{\text{reg}}^*$ .

**Proposition 3.2** (Generalized Harish-Chandra expansion; Theorem 5.9.4 of [GV88]). *Let  $\varepsilon > 0$ . There exist constants  $C > 0$  and  $s \geq 0$  such that for all  $\lambda \in \mathfrak{a}_{\text{reg}}^*$  and for all  $H \in \bar{\mathfrak{a}}_+$  such that  $\beta_L(H) \geq \varepsilon$ , we have*

$$\left| e^{\rho(H)} \varphi_\lambda(e^H) - e^{\rho_L(H)} \theta_L(H, \lambda) \right| \leq C(1 + \|\lambda\|)^s (1 + \|H\|)^s e^{-2\beta_L(H)}.$$

We shall discuss an extension of this result to a small tube about the tempered subspace  $\mathfrak{a}^*$  in Proposition 9.3. In case  $L = Z_K(\mathfrak{a})A$ , so that  $L$  is the unique minimal standard Levi, we have that  $\rho_L = 0$ , and  $\varphi_\lambda^L(e^H) = e^{\lambda(H)}$ . The formula then reduces to the classical Harish-Chandra expansion.

**3.6. Maximally singular and extremally singular elements.** An element  $H_0 \in \bar{\mathfrak{a}}_+$  is called called *singular* if  $H_0 \in \bar{\mathfrak{a}}_+ \setminus \mathfrak{a}_+$ . Non-zero elements lying on the one-dimensional faces of the polyhedral cone  $\bar{\mathfrak{a}}_+$  are the most singular among all non-zero elements and shall be called *maximally singular*. The fundamental coweights give a conical basis for  $\bar{\mathfrak{a}}_+$ , so the maximally singular elements are simply the positive multiples of fundamental coweights. The fundamental coweights correspond bijectively with the nodes of the Dynkin diagram.

Among the maximally singular elements in  $\bar{\mathfrak{a}}_+$  we would now like to isolate a subclass which enjoys certain extremal properties. We will only be able to find such a nice class in types  $A_n$ ,  $B_n$ ,  $C_n$ ,  $D_n$ , and  $E_7$ , so for the rest of this subsection we shall assume that  $\Phi_{\text{red}}$  is of one of these types. We say that  $H_0 \in \bar{\mathfrak{a}}_+$  is *extremally singular* (or *extremal*) if  $H_0$  is equal to a positive multiple of an *extremal* fundamental coweight. The latter are defined via the following table, which identifies the node (or nodes) of the Dynkin diagram of  $\Phi_{\text{red}}$  which corresponds to the extremal fundamental coweights:

type of $\Phi_{\text{red}}$	extremal nodes $\bullet$	type of $\Phi_{M,\text{red}}$
$A_n$ ( $n \geq 1$ )	$\bullet - \circ - \cdots - \circ - \bullet$	$A_{n-1}$
$B_n$ ( $n \geq 2$ )	$\bullet - \circ - \cdots - \circ - \circ \circ - \circ$	$B_{n-1}$
$C_n$ ( $n \geq 3$ )	$\bullet - \circ - \cdots - \circ - \circ \circ - \circ$	$C_{n-1}$
$D_n$ ( $n \geq 4$ )	$\bullet - \circ - \cdots - \circ - \circ \circ - \circ$	$D_{n-1}$
$E_7$	$\circ - \circ - \circ - \circ - \circ - \circ - \bullet$	$E_6$

TABLE 1. Darkened nodes correspond to extremal fundamental coweights

For the infinite families  $A_n, B_n, C_n, D_n$ , the extremal nodes are the ones that you can remove from the Dynkin diagram to obtain a Dynkin diagram in the same infinite family but of rank one less (with the conventions  $B_1 = A_1, C_2 = B_2, D_3 = A_3$ ).

**3.7. Benjamini–Schramm convergence.** A lattice  $\Gamma < G$  is a discrete subgroup of finite covolume. The lattice  $\Gamma$  is said to be *uniform* if  $Y$  is compact and *irreducible* if the projection of  $\Gamma$  onto each simple factor is dense. We will generally assume that  $\Gamma$  is uniform, irreducible, and torsion-free.

The injectivity radius of a point  $y \in Y$ , denoted  $\text{InjRad}_Y(y)$ , is the supremum of all  $r$  such that the ball of radius  $r$  centered at any lift  $\tilde{y} \in X$  of  $y$  maps injectively to  $Y$  under the canonical projection  $X \rightarrow Y$ . The injectivity radius of  $Y$ , denoted  $\text{InjRad}(Y)$ , is the infimum over all  $y \in Y$  of  $\text{InjRad}_Y(y)$ . We define

$$Y_{\leq R} := \{y \in Y : \text{InjRad}_Y(y) \leq R\}.$$

Suppose  $Y_n$  is a sequence of locally symmetric spaces obtained from quotienting  $X$  by a sequence of torsion-free lattices. We say that the sequence  $Y_n$  *Benjamini–Schramm converges* to  $X$  if, for every  $R > 0$ , we have

$$\frac{\text{vol}((Y_n)_{\leq R})}{\text{vol}(Y_n)} \rightarrow 0.$$

We say that the sequence  $Y_n$  is *uniformly discrete* if there is a universal non-zero lower bound on  $\text{InjRad}(Y_n)$ .

**3.8. Plancherel convergence.** Let  $\widehat{G}$  denote the unitary dual of  $G$ , consisting of isomorphism classes of irreducible unitary representations of  $G$ , and endowed with the Fell topology. Suppose  $\Gamma$  is a cocompact lattice in  $G$ . Then

$$(25) \quad L^2(\Gamma \backslash G) = \bigoplus_{\pi \in \widehat{G}} m(\pi, \Gamma) \pi,$$

where the multiplicity  $m(\pi, \Gamma) = \dim \text{Hom}_G(\pi, L^2(\Gamma \backslash G))$  is finite, and  $m(\pi, \Gamma) = 0$  for all but countably many  $\pi$ . We define the spectral measure on  $\widehat{G}$  relative to  $\Gamma$  as

$$\mu_\Gamma := \frac{1}{\text{vol}(\Gamma \backslash G)} \sum_{\pi \in \widehat{G}} m(\pi, \Gamma) \delta_\pi.$$

An important component of the proof of Theorem 1.1 is the limiting behavior of the spectral measure  $\mu_\Gamma$  when  $Y = \Gamma \backslash X$  converges Benjamini–Schramm to  $X$ .

Recall that we have fixed a Haar measure  $dg$  on  $G$ . If  $f \in C_c^\infty(G)$  and  $\pi$  is an irreducible unitary representation of  $G$ , we define the trace class operator  $\pi(f) := \int_G f(g) \pi(g) dg$ . The Plancherel measure  $d\mu_{\text{Pl}}$  is the unique Radon measure on  $\widehat{G}$  verifying the inversion formula  $f(e) = \int_{\pi \in \widehat{G}} \text{tr}(\pi(f)) d\mu_{\text{Pl}}(\pi)$ .

As our main interest is  $L^2(\Gamma \backslash X)$  rather than  $L^2(\Gamma \backslash G)$ , as in (25), we shall restrict our attention to irreducible unitary *spherical* representations  $\pi$  of  $G$ . An irreducible representation is said to be spherical if the space of  $K$ -invariant vectors is non-zero, in which case it is one-dimensional. Any such representation can be realized as the unique spherical subquotient  $\pi_\lambda$  of the (unitarily normalized) principal series representation  $\text{Ind}_P^G \chi_\lambda$ , where  $\lambda \in \mathfrak{a}_\mathbb{C}^*$  and  $\chi_\lambda$  is the character  $\exp(\lambda(\mathcal{H}(p)))$  of the minimal parabolic  $P$ . Since  $\pi_\lambda \simeq \pi_{\lambda'}$  if and only if there is  $w \in W$  with  $\lambda = w\lambda'$ , the map sending  $\pi_\lambda$  to  $\lambda$  descends to an injective map

from the spherical unitary dual  $\widehat{G}^{\text{sph}}$  to  $\mathfrak{a}_{\mathbb{C}}^*/W$ . Let the image of this injection be denoted  $\mathfrak{a}_{\text{un}}^*/W$ ; it contains all of  $\mathfrak{a}^*/W$ . By comparison, for any  $\lambda \in \mathfrak{a}_{\mathbb{C}}^*$ , the matrix coefficient (relative to the inner product given by integration over  $K$ , which is unitary only for  $\lambda \in \mathfrak{a}^*/W$ ) of any unit vector in  $\pi_\lambda^K$  recovers the Harish-Chandra integral expression for the spherical function in (16). Without abuse, we may then speak of  $\lambda$  as the *spectral parameter* for  $\pi_\lambda$ . In this parametrization, the restriction of the Plancherel measure to the unitary spherical representations has density function equal to  $|c(\lambda)|^{-2}$  on  $\mathfrak{a}^*/W$  and is identically zero outside of this locus. Moreover, if  $\lambda \in \mathfrak{a}_{\text{un}}^*/W$  and  $f \in C_c^\infty(G//K)$ , then  $\text{tr } \pi_\lambda(f) = \widehat{f}(\lambda)$ , the Harish-Chandra spherical transform of  $f$  evaluated at  $\lambda$ , as defined in (17).

We have the following result [Dei18; ABB+17].

**Theorem 3.3.** *Suppose  $\Gamma_n$  is a sequence of cocompact, uniformly discrete lattices in  $G$ . Then the following are equivalent:*

- (1) *The sequence of locally symmetric spaces  $\Gamma_n \backslash X$  Benjamini–Schramm converges to  $X$ .*
- (2) *For every  $f \in C_c^\infty(G)$*

$$(26) \quad \int_{\widehat{G}} \text{tr } \pi(f) d\mu_{\Gamma_n}(\pi) \rightarrow \int_{\widehat{G}} \text{tr } \pi(f) d\mu_{\text{Pl}}(\pi).$$

For  $\lambda \in \mathfrak{a}_{\text{un}}^*$  let  $L^2(\Gamma \backslash X)_\lambda$  denote the  $K$ -invariant subspace of the  $\pi_\lambda$ -isotypic component of  $L^2(\Gamma \backslash G)$ . Taking  $K$ -invariants in (25), we obtain an orthonormal basis  $\mathcal{B}_\Gamma = \{\psi_j\}_{j \geq 0}$  of  $L^2(\Gamma \backslash X)$  such that  $\psi_j \in L^2(\Gamma \backslash X)_{\lambda_j}$ . The functions  $\psi_j$  are the Maass forms introduced in §1.2 and  $\lambda_j$  is their spectral parameter. For a  $W$ -invariant subset  $\Omega \subset \mathfrak{a}^*/W$  recall from the statement of Theorem 1.1 that  $N(\Omega, \Gamma) = \sum_{\lambda_j \in \Omega} \dim L^2(\Gamma \backslash X)_{\lambda_j}$ .

The arguments of Section 6 in [Pet23b], with slight modifications, imply the following result.

**Proposition 3.4.** *Suppose  $\Gamma_n$  is a sequence of cocompact uniformly discrete torsion-free lattices in  $G$  such that the corresponding sequence of locally symmetric spaces Benjamini–Schramm converges to  $X$ . Let  $\Omega \subset \mathfrak{a}^*/W$  be a bounded measurable subset such that  $\mu_{\text{Pl}}(\partial\Omega) = 0$ . Then*

$$\left| \frac{N(\Omega, \Gamma_n)}{\text{vol}(Y_n)} - \mu_{\text{Pl}}(\Omega) \right| \rightarrow 0$$

as  $n \rightarrow \infty$ .

For our purposes, a non-zero lower bound on  $\frac{N(\Omega, \Gamma_n)}{\text{vol}(Y_n)}$  ultimately suffices. Thus, as in the statement of Theorem 1.1, it suffices to assume that  $\Omega$  is compact with non-empty interior, as in this case it contains an open ball.

In the deduction of Proposition 3.4 from Theorem 3.3, the main difficulty is that the indicator function of  $\Omega$  is not of the form  $\widehat{f}$  for some  $f \in C_c^\infty(G//K)$ . The latter class of functions may be identified, using the Harish-Chandra Paley–Wiener theorem, with the  $W$ -invariant functions in the Paley–Wiener space  $\mathcal{PW}(\mathfrak{a}_{\mathbb{C}}^*)$ . All such functions restricted to  $\mathfrak{a}_{\text{un}}^*/W$  vanish at infinity and thus we may use the Stone–Weierstrass theorem (together with Urysohn’s lemma) to approximate  $\mathbf{1}_\Omega$  on  $\mathfrak{a}_{\text{un}}^*/W$  by elements in  $\mathcal{PW}(\mathfrak{a}_{\mathbb{C}}^*)^W$ . From here the arguments in [Pet23b], in particular the proofs of Proposition 6.5 and Theorem 1.9, can be repeated.

**3.9. Spectral gap.** If  $(\pi, V_\pi)$  is a (not necessarily irreducible) separable unitary representation of  $G$ , then we define its integrability exponent by

$$q(\pi) := \inf\{q \geq 2 : \langle \pi(g)v_1, v_2 \rangle \in L^q(G) \text{ for } v_1, v_2 \text{ in a dense subspace of } V_\pi\}.$$

We say that  $(\pi, V_\pi)$  has a spectral gap if  $q(\pi) < \infty$ , and that a family of unitary representations  $\{(\pi_n, V_{\pi_n})\}$  has a uniform spectral gap if we can find a uniform upper bound on  $q(\pi_n)$ . We call the representation *tempered* if  $q(\pi) = 2$ .

#### 4. SPECTRAL ESTIMATE

In this section, we establish an upper bound on the spectral sums appearing in (1) by averaging over an expanding spherical shell. As in the classical proof of quantum ergodicity, the insertion of this time dependent averaging operator, or wave propagator, will allow us in later sections to use their ergodic properties in the presence of a spectral gap.

We now drop the dependence on  $n$  in the subscripts, writing  $\Gamma < G$  for a cocompact lattice in  $G$ ,  $Y = \Gamma \backslash X$ , and  $\psi_j$  for the orthonormal basis of Maass forms on  $Y$ .

**4.1. Averaging set.** We fix a simple factor  $G_1$  and let  $\mathcal{G}_2 = G_2 \times \cdots \times G_s$  denote the product of the remaining simple factors (if any). Let  $K_1$  (resp.  $\mathcal{K}_2$ ) and  $A_1$  (resp.  $\mathcal{A}_2$ ) denote the images of  $K$  and  $A$  inside  $G_1$  (resp.  $\mathcal{G}_2$ ). We decompose  $\mathfrak{a}$  as  $\mathfrak{a}_1 \oplus \text{Lie}(\mathcal{A}_2)$ , where  $\mathfrak{a}_1 = \text{Lie}(A_1)$ . For  $r > 0$  and  $H \in \mathfrak{a}_1$  let  $B_{\mathfrak{a}_1}(H, r)$  denote the Euclidean ball in  $\mathfrak{a}_1$  of radius  $r$  and centered at  $H$ . Similarly,  $\mathcal{B}_2(0, r)$  denotes the Euclidean ball in  $\text{Lie}(\mathcal{A}_2)$  of radius  $r$ , centered at 0.

Let  $H_0 \in \bar{\mathfrak{a}}_{1,+}$  be non-zero and  $\epsilon_0, t > 0$ . In practice,  $t$  will be large and tending to  $\infty$ ,  $\epsilon_0$  will be sufficiently small but fixed. We shall sometimes refer to  $H_0$  as the *directing element*; it will be fixed and all implied constants will depend on it. We let

$$(27) \quad S_t = K_1 \exp(B_{\mathfrak{a}_1}(tH_0, \epsilon_0))K_1 \subset G_1 \quad \text{and} \quad B = \mathcal{K}_2 \exp(\mathcal{B}_2(0, \epsilon_0))\mathcal{K}_2 \subset \mathcal{G}_2.$$

We will propagate eigenfunctions along the subset

$$(28) \quad E_t = S_t \times B \subset G$$

using the right-regular representation  $\varrho_{\Gamma \backslash G}$  of  $G$  on  $L^2(\Gamma \backslash G)$ . Note that, for any function  $f \in L^1(G)$ , the adjoint of  $\varrho_{\Gamma \backslash G}(f)$  is given by  $\varrho_{\Gamma \backslash G}(f^\vee)$ , where  $f^\vee(g) := \overline{f(g^{-1})}$ . We may then define an operator  $U_t$  and its adjoint  $U_t^*$  by

$$U_t := \varrho_{\Gamma \backslash G}(e^{-t\rho(H_0)} \mathbf{1}_{E_t}) \quad \text{and} \quad U_t^* := \varrho_{\Gamma \backslash G}(e^{-t\rho(H_0)} \mathbf{1}_{E_t^{-1}}).$$

Note that, for  $t$  sufficiently large,

$$\text{vol}(E_t) \asymp \int_{B_{\mathfrak{a}_1}(tH_0, \epsilon_0) \cap \mathfrak{a}_{1,+}} J(H) dH \asymp e^{2t\rho(H_0)},$$

where we have used (11) and (12), and where  $J(H)$  here denotes the radial volume factor for  $G_1$ . It follows that the normalization factor of  $e^{-t\rho(H_0)}$  in the definition of  $U_t$  is essentially the square-root of the volume of  $E_t$ . Observe furthermore that the propagation takes place solely within  $G_1$ , as the ball  $B \subset \mathcal{G}_2$  is independent of the parameter  $t$ . This will eventually allow us to isolate the conditions necessary to successfully analyze the ergodic properties of  $U_t$  to the privileged factor  $G_1$ , as is reflected in the statement of Theorem 1.1.

We have elected to suppress the dependency in the notation for the sets  $E_t, S_t, B$ , as well as for the operator  $U_t$ , on  $H_0$  and  $\epsilon_0$ . While this lightens the notational load, we remark that

in some places, such as in Theorems 4.1 and 5.1, it will be necessary to take  $\epsilon_0$  sufficiently small. More importantly, it will be crucial in Section 6, wherein we prove Theorem 2.2, that  $H_0$  be chosen extremely singular.

Some comments are in order on the relation between the sets  $E_t$  and the sets used for similar purposes elsewhere in the literature. The averaging operators used in the work of Le Masson–Sahlsten [LMS17] were defined by suitably normalized hyperbolic balls. In the higher rank settings of Brumley–Matz [BM23] and Peterson [Pet23b], “polytopal ball” averaging operators were used instead. The operator  $U_t$  introduced above dispenses with the polytopes, but preserves their core features, by averaging only in a small neighborhood of the directing element  $H_0$ . The operator  $U_t$ , and the choice of  $H_0$  implicit in its definition, plays a similar role for the real Lie group  $G$  that the normalized Hecke operator

$$q^{-\langle \mu, \rho \rangle} \mathbb{1}_{\mathbf{G}(\mathcal{O})\varpi^\mu \mathbf{G}(\mathcal{O})}$$

associated with a choice of dominant cocharacter  $\mu$ , plays on an algebraic group  $\mathbf{G}$  over a non-archimedean local field (with ring of integers  $\mathcal{O}$ , residue field of order  $q$ , and uniformizer  $\varpi$ ). Note that, similarly to the archimedean setting,  $\text{vol}(\mathbf{G}(\mathcal{O})\varpi^{t\mu} \mathbf{G}(\mathcal{O})) \asymp q^{\langle t\mu, 2\rho \rangle}$ .

Ultimately, the fundamental problem, which we have already emphasized in §2.2, is to choose a directing element  $H_0$  for which the corresponding intersection volumes are minimized, as in Section 6. By contrast, the results we prove in this section are valid more generally, with significantly fewer constraints on the group  $G$  and the directing element  $H_0$ .

**4.2. Main spectral theorem and reduction to a local integral.** Let  $a$  be a bounded measurable function on  $Y$ . When viewed as a right- $K$ -invariant function on  $\Gamma \backslash G$ , the function  $a$  determines a multiplication operator on  $L^2(\Gamma \backslash G)$ . We consider the time average

$$(29) \quad \mathbf{A}(\tau) = \frac{1}{\tau} \int_\tau^{2\tau} U_t a U_t^* dt.$$

In line with our convention for  $U_t$ , we have suppressed the dependency on the parameter  $\epsilon_0$ , as well as the directing element  $H_0$ , from the notation  $\mathbf{A}(\tau)$ .

The aim of this section is to prove the following result.

**Theorem 4.1** (Spectral estimate). *Let  $G$  be a product of non-compact simple real Lie groups, and  $\Gamma < G$  an irreducible lattice. As in §4.1, we fix a simple factor  $G_1$  of  $G$  and let  $H_0 \in \bar{\mathfrak{a}}_{1,+}$  be non-zero. Then there exists a finite  $W$ -stable set of hyperplanes  $\{P_i\}$  in  $\mathfrak{a}^*$ , depending on  $H_0$ , such that the following holds. Let  $\Omega \subset \mathfrak{a}^* \setminus \cup_i P_i$  be compact and  $W$ -invariant. There are constants  $c, \tau_0, \epsilon_0 > 0$ , depending on  $\Omega$ , such that for all  $\tau \geq \tau_0$  and all  $a \in L^\infty(Y)$  we have*

$$\sum_{j: \lambda_j \in \Omega} |\langle a\psi_j, \psi_j \rangle|^2 \leq c \sum_{j: \lambda_j \in \Omega} \left| \langle \mathbf{A}(\tau)\psi_j, \psi_j \rangle \right|^2.$$

Here,  $\mathbf{A}(\tau)$  is defined relative to the parameters  $\epsilon_0 > 0$  and  $H_0$ .

We begin by reducing Theorem 4.1 to a purely local statement (independent of  $\Gamma$ ), involving only  $G_1$ . Using the notation introduced in §4.1, we have  $\mathfrak{a}^* = \mathfrak{a}_1^* \oplus \text{Lie}(\mathcal{A}_2)^*$ . We may decompose  $\lambda \in \mathfrak{a}^*$  as  $\lambda = \lambda_1 + \lambda_2$  according to this decomposition. Using the fact that  $U_t^*$  acts on Maass forms  $\psi_\lambda$  of spectral parameter  $\lambda$  by the scalar

$$e^{-t\rho(H_0)} \widehat{\mathbb{1}_{E_t^{-1}}}(-\lambda) = e^{-t\rho(H_0)} \widehat{\mathbb{1}_{E_t}}(\lambda) = e^{-t\rho(H_0)} \widehat{\mathbb{1}_{S_t}}(\lambda_1) \widehat{\mathbb{1}_B}(\lambda_2),$$

we find

$$\begin{aligned}
\langle \mathbf{A}(\tau)\psi_\lambda, \psi_\lambda \rangle &= \frac{1}{\tau} \int_\tau^{2\tau} \langle U_t a U_t^* \psi_\lambda, \psi_\lambda \rangle dt \\
&= \frac{1}{\tau} \int_\tau^{2\tau} \langle a U_t^* \psi_\lambda, U_t^* \psi_\lambda \rangle dt \\
(30) \quad &= |\widehat{\mathbb{1}_B}(\lambda_2)|^2 \left( \frac{1}{\tau} \int_\tau^{2\tau} e^{-2t\rho(H_0)} |\widehat{\mathbb{1}_{S_t}}(\lambda_1)|^2 dt \right) \langle a \psi_\lambda, \psi_\lambda \rangle.
\end{aligned}$$

Here,  $\widehat{\mathbb{1}_E}$  denotes the Harish-Chandra transform (17), for the group  $G_1$  or  $\mathcal{G}_2$  as appropriate, of the characteristic function of a subset  $E$ .

Let  $\Omega_2 \subset \text{Lie}(\mathcal{A}_2)^*$  be compact and invariant under the Weyl group for  $\mathcal{G}_2$ . An elementary argument using compactness, to be given below, shows that  $\widehat{\mathbb{1}_B}(\lambda_2)$  is bounded away from zero, uniformly for  $\lambda_2 \in \Omega_2$ .

**Lemma 4.2.** *Let  $G$  be a product of non-compact simple real Lie groups. Let  $\Omega \subset \mathfrak{a}^*$  be compact and  $W$ -invariant. Then there are constants  $\epsilon_0, c > 0$  such that the characteristic function  $\mathbb{1}_B$  of  $B = K \exp(B_{\mathfrak{a}}(0, \epsilon_0))K$  satisfies  $|\widehat{\mathbb{1}_B}(\lambda)| \geq c$  uniformly for  $\lambda \in \Omega$ .*

*Proof.* As usual, let  $\varphi_\lambda$  denote the spherical function on  $G$ . For all  $\lambda \in \mathfrak{a}_{\mathbb{C}}^*$ , we have  $\varphi_\lambda(e) = 1$ ; moreover, the map

$$\text{Re } \varphi : \mathfrak{a} \times \mathfrak{a}^* \rightarrow \mathbb{R}, \quad (H, \lambda) \mapsto \text{Re } \varphi_{-\lambda}(e^H)$$

is continuous. Thus, for any  $\lambda \in \mathfrak{a}^*$  we may find a neighborhood  $U_\lambda \times V_\lambda \subset \mathfrak{a} \times \mathfrak{a}^*$  of  $(0, \lambda)$  on which  $\text{Re } \varphi > 1/2$ , say. By the compactness of  $\Omega$  there exists a finite subcover  $\{V_{\lambda_j}\}_{j=1}^N$  of  $\{V_\lambda : \lambda \in \Omega\}$ ; let  $U = \bigcap_{j=1}^N U_{\lambda_j}$ . Then  $\text{Re } \varphi_{-\lambda}(e^H) > 1/2$  for all  $H \in U$  and  $\lambda \in \Omega$ . We can therefore choose  $\epsilon_0 > 0$  sufficiently small so that

$$\text{Re} \left( \widehat{\mathbb{1}_B}(\lambda) \right) = b_G \int_{B_{\mathfrak{a}}(0, \epsilon_0) \cap \mathfrak{a}_+} \text{Re } \varphi_{-\lambda}(e^H) J(H) dH > \frac{b_G}{2} \int_{B_{\mathfrak{a}}(0, \epsilon_0) \cap \mathfrak{a}_+} J(H) dH > 0,$$

for all  $\lambda \in \Omega$ , as desired.  $\square$

We are therefore reduced to proving uniform lower bounds for the integral over  $t$  in (30). We note that Lemma 4.2 imposes no further conditions on  $\Omega_2$ , beyond compactness and Weyl group invariance. In view of the factorization of the scalar factors in (30) according to the components  $\lambda_1$  and  $\lambda_2$ , this implies that the hyperplanes  $P_i \subset \mathfrak{a}_1^* \oplus \text{Lie}(\mathcal{A}_2)^*$  that  $\Omega$  must avoid in Theorem 4.1 can, and will, be taken to contain  $\text{Lie}(\mathcal{A}_2)^*$ . Their exact nature is described in the following result (with  $G$  playing the role of the privileged factor  $G_1$ ), which will then be enough to prove (a more precise form of) Theorem 4.1.

**Proposition 4.3.** *Let  $G$  be a non-compact simple Lie group. Let  $H_0 \in \bar{\mathfrak{a}}_+$  be non-zero, with centralizer  $M$ . Fix a compact and  $W$ -invariant subset*

$$(31) \quad \Omega \subset \mathfrak{a}^* \setminus \mathfrak{a}_{\text{bad}}^*, \quad \mathfrak{a}_{\text{bad}}^* := \bigcup_{w \notin W_M} W \{ \lambda \in \mathfrak{a}^* : \lambda(H_0 - wH_0) = 0 \}.$$

*Then there is  $\epsilon_0 > 0$  (implicit in the definition of  $S_t$ ) as well as constants  $c, \tau_0 > 0$  such that*

$$\frac{1}{\tau} \int_\tau^{2\tau} e^{-2t\rho(H_0)} |\widehat{\mathbb{1}_{S_t}}(\lambda)|^2 dt > c$$

*for all  $\lambda \in \Omega$  and all  $\tau \geq \tau_0$ .*

The basic strategy of the proof of Proposition 4.3 is to use the generalized Harish-Chandra expansion (Proposition 3.2) of the spherical function  $\varphi_\lambda$  along the direction determined by  $H_0$  to reduce the estimate to a weighted sum of complex exponentials. For the off-diagonal terms, one then needs to provide a uniform bound in  $\tau$  of the quantity

$$\max_{\substack{w, w' \in W/W_M \\ w \neq w'}} \left| \int_{\tau}^{2\tau} e^{-it\lambda(wH_0 - w'H_0)} dt \right|.$$

Such a uniform bound will reflect the oscillation of the integrand, provided the phases  $\lambda(wH_0 - w'H_0)$  are bounded away from zero. The latter condition is ensured by the hypothesis on  $\Omega$ .

**4.3. Reduction to complex exponentials.** Recall that  $M$  is the centralizer of  $H_0$  in  $G$ . Using the Harish-Chandra expansion along  $M$  for the spherical function  $\varphi_{-\lambda}$ , we now reduce the proof of Proposition 4.3 to a corresponding lower bound, stated in Proposition 4.6 below, in which the variation in the  $t$ -parameter is expressed solely through complex exponentials. The key input is the next lemma, which uses the generalized Harish-Chandra expansion, as formulated in Proposition 3.2, as a critical ingredient.

**Lemma 4.4.** *Let  $G$  be a non-compact simple Lie group. Let  $\Omega \subset \mathfrak{a}_{\text{reg}}^*$  be compact and  $W$ -invariant. Let  $H_0 \in \bar{\mathfrak{a}}_+$  with centralizer  $M$ . Put  $C_M = (b_G/b_M)2^{-\ell}$ , where  $\ell = \sum_{\alpha \in \Phi^+ \setminus \Phi_M^+} m_\alpha$  and  $b_L$  is defined in (14). Let  $\varepsilon := \min_{\alpha \in \Phi^+ \setminus \Phi_M^+} \alpha(H_0) > 0$ .*

*For any  $\epsilon_0 > 0$  (implicit in the definition of  $S_t$ ) there is  $t_0 > 0$  such that for all  $\lambda \in \Omega$  and all  $t \geq t_0$  we have*

$$e^{-t\rho(H_0)} \widehat{\mathbf{1}_{S_t}}(\lambda) = C_M \sum_{w \in W/W_M} c^M(w\lambda) \widehat{\mathbf{1}_{B^M}}(w\lambda) e^{-itw\lambda(H_0)} + O(\epsilon_0 + e^{-\varepsilon t}),$$

where  $B^M = K_M \exp(B_{\mathfrak{a}}(0, \epsilon_0)) K_M \subset M$ .

*Proof.* It follows from the definition (17) that

$$e^{-t\rho(H_0)} \widehat{\mathbf{1}_{S_t}}(\lambda) = b_G e^{-t\rho(H_0)} \int_{B_{\mathfrak{a}}(tH_0, \epsilon_0)_+} \varphi_{-\lambda}(e^H) J(H) dH,$$

where  $B_{\mathfrak{a}}(tH_0, \epsilon_0)_+ = B_{\mathfrak{a}}(tH_0, \epsilon_0) \cap \mathfrak{a}_+$ . Since  $\Omega$  avoids the singular locus, we may apply Proposition 3.2 to  $\varphi_{-\lambda}$ , with  $L = M$ , to get

$$(32) \quad e^{-t\rho(H_0)} \widehat{\mathbf{1}_{S_t}}(\lambda) = \left[ \sum_{w \in W/W_M} c^M(w\lambda) I^M(w\lambda) \right] + O(\mathcal{E}_t),$$

the implied constant depending only on  $\Omega$ , where

$$I^M(\lambda) = b_G \int_{B_{\mathfrak{a}}(tH_0, \epsilon_0)_+} \left( e^{-t\rho(H_0) - \rho^M(H)} J^M(H) \right) \varphi_{-\lambda}^M(e^H) J_M(H) dH,$$

with  $\rho^M(H) = \rho(H) - \rho_M(H)$  and the factor  $J_M(H)$  defined in (15), and

$$\mathcal{E}_t = \int_{B_{\mathfrak{a}}(tH_0, \epsilon_0)_+} \left( e^{-t\rho(H_0) - \rho(H)} J(H) \right) e^{-2\beta_M(H)} (1 + \|H\|)^s dH,$$

with  $\beta_M(H)$  defined in (24).

We now let  $S_t^M = e^{tH_0}B^M = K_M \exp(B_{\mathfrak{a}}(tH_0, \epsilon_0))K_M$  be the ball of radius  $\epsilon_0$  about  $e^{tH_0}$  in  $M$ . We claim that for any  $\epsilon_0 > 0$  there is  $t_0 > 0$  such that for all  $t \geq t_0$ , we have

$$(33) \quad I^M(\lambda) = C_M \widehat{\mathbb{1}_{S_t^M}}(\lambda) + O(\epsilon_0) \quad \text{and} \quad \mathcal{E}_t = O(e^{-\varepsilon t}).$$

Inserting this into (32), we recover the statement of the lemma. Indeed,  $\max_{\lambda \in \Omega} |c^M(\lambda)| \ll 1$ , since  $\Omega$  is bounded away from the polar hyperplanes of  $c^M$ , and

$$\widehat{\mathbb{1}_{S_t^M}}(\lambda) = \int_M \mathbb{1}_{B^M}(e^{-tH_0}m) \varphi_{-\lambda}^M(m) dm = \int_M \mathbb{1}_{B^M}(m) \varphi_{-\lambda}^M(e^{tH_0}m) dm = e^{-it\lambda(H_0)} \widehat{\mathbb{1}_{B^M}}(\lambda).$$

In the last equality, we have used the fact, which follows from (16), that  $\varphi_{\lambda}^M(e^Z m) = e^{i\lambda(Z)} \varphi_{\lambda}^M(m)$  whenever  $Z \in \mathfrak{a}$  centralizes  $M$ .

For the main term identity in (33), we begin by observing that the parenthetical expression in the definition of  $I^M(\lambda)$  is asymptotically constant on the support of the integral:

$$(34) \quad e^{-t\rho(H_0) - \rho^M(H)} J^M(H) = 2^{-\ell}(1 + O(\epsilon_0)), \quad \forall H \in B_{\mathfrak{a}}(tH_0, \epsilon_0), \quad \forall t > 0.$$

To see this, we use  $\alpha(H_0) > 0$  for  $\alpha \in \Phi^+ \setminus \Phi_M^+$ , to show that, for such  $H$ :

$$J^M(H) = 2^{-\ell} \left( \prod_{\alpha \in \Phi^+ \setminus \Phi_M^+} (e^{\alpha(H)} - e^{-\alpha(H)})^{m_{\alpha}} \right) = 2^{-\ell} e^{2\rho^M(H)} + O(e^{t(1-\beta_M(H_0))2\rho^M(H_0)}).$$

Using this and  $e^{\rho^M(H)} = e^{t\rho^M(H_0)} e^{\rho^M(H-tH_0)} = e^{t\rho(H_0)}(1 + O(\epsilon_0))$ , we obtain the claim (34), which may then be inserted into  $I^M(\lambda)$  to yield

$$I^M(\lambda) = C_M b_M \int_{B_{\mathfrak{a}}(tH_0, \epsilon_0)_+} \varphi_{-w\lambda}^M(e^H) J_M(H) dH + O(\epsilon_0).$$

We fix  $t_0$  large enough so that  $B_{\mathfrak{a}}(tH_0, \epsilon_0)_+ = B_{\mathfrak{a}}(tH_0, \epsilon_0) \cap \mathfrak{a}_{M,+}$  for all  $t > t_0$ , where  $\mathfrak{a}_{M,+}$  is defined in (13). In particular, the union  $\bigcup_{w \in W_M} wB_{\mathfrak{a}}(tH_0, \epsilon_0)_+$  is disjoint, with closure  $\overline{B}_{\mathfrak{a}}(tH_0, \epsilon_0)$ . Since  $J_M(wH) = J_M(H)$ ,  $\varphi_{\lambda}^M(e^{wH}) = \varphi_{\lambda}^M(e^H)$  for  $w \in W_M$ , the integration formula (14) implies that

$$b_M \int_{B_{\mathfrak{a}}(tH_0, \epsilon_0)_+} \varphi_{-\lambda}^M(e^H) J_M(H) dH = \widehat{\mathbb{1}_{S_t^M}}(\lambda)$$

as required.

To estimate the error term  $\mathcal{E}_t$ , we first apply (12) to deduce that  $e^{-t\rho(H_0) - \rho(H)} J(H) \ll 1$  on  $B_{\mathfrak{a}}(tH_0, \epsilon_0)_+$ . Moreover,

$$\beta_M(tH_0) = \min_{\alpha \in \Phi^+ \setminus \Phi_M^+} \alpha(tH_0) = 2\varepsilon t.$$

Given  $\epsilon_0 > 0$  there exists  $t_0 > 0$  such that  $\beta_M(H) \geq (3\varepsilon/2)t$  for all  $t \geq t_0$  and  $H \in B_{\mathfrak{a}}(tH_0, \epsilon_0)$ . We deduce

$$\mathcal{E}_t = O \left( e^{-(3\varepsilon/2)t} \int_{B_{\mathfrak{a}}(tH_0, \epsilon_0)_+} (1 + \|H\|)^s dH \right) = O(e^{-\varepsilon t}),$$

by the polynomial growth in  $t$  of the integral.  $\square$

Lemma 4.4 requires that the compact  $\Omega$  avoid the singular locus. Proposition 4.3, on the other hand, requires that  $\Omega$  avoid  $\mathfrak{a}_{\text{bad}}^*$ , a set of hyperplanes depending on  $H_0$ . The following lemma makes use of our assumption that  $H_0$  is non-zero to show an inclusion of the former in the latter.

**Lemma 4.5.** *Let  $G$  be a non-compact simple Lie group. Let  $H_0 \in \bar{\mathfrak{a}}_+$  be non-zero. Then  $\mathfrak{a}_{\text{sing}}^* \subseteq \mathfrak{a}_{\text{bad}}^*$ .*

*Proof.* We must show that

$$(35) \quad \bigcup_{w, w' \in W} \mathbb{R}.(wH_0 - w'H_0) \supseteq \bigcup_{\alpha \in \Phi} \mathbb{R}.\alpha^\vee.$$

Since  $W_M$  consists precisely of those elements of  $W$  which fix  $H_0$ , we may rewrite the LHS of (35) as

$$(36) \quad \begin{aligned} \bigcup_{w, w' \in W} \mathbb{R}.w(H_0 - w^{-1}w'H_0) &= \bigcup_{w \in W} \bigcup_{w' \in W} \mathbb{R}.w(H_0 - w'H_0) \\ &= \bigcup_{w' \in W/W_M} W\mathbb{R}.(H_0 - w'H_0). \end{aligned}$$

For  $\alpha \in \Phi$  we let  $s_\alpha \in W$  denote the orthogonal reflection in  $\mathfrak{a}$  across  $(\alpha^\vee)^\perp$ . In particular,

$$(37) \quad H_0 - s_\alpha H_0 = 2 \frac{\langle H_0, \alpha^\vee \rangle}{\langle \alpha^\vee, \alpha^\vee \rangle} \alpha^\vee.$$

Note that the subset of all long roots in  $\Phi$  spans  $\mathfrak{a}^*$ , as does the subset of all short roots. Since  $H_0 \neq 0$  there must therefore be a long root  $\beta_1$  and a short root  $\beta_2$  such that  $\langle H_0, \beta_1^\vee \rangle \neq 0$  and  $\langle H_0, \beta_2^\vee \rangle \neq 0$ . Therefore  $\beta_1$  and  $\beta_2$  are contained in the LHS of (35). Since  $W$  acts transitively on the subset of all long roots in  $\Phi$  and also on the subset of all short roots, all of  $\Phi$  is therefore contained in the left hand side of (35).  $\square$

**4.4. Proof of Proposition 4.3.** Our goal is to prove Proposition 4.3 which, as we have seen in §4.2, implies Theorem 4.1. From Lemmas 4.4 and 4.5, it suffices to show the following result. We explain this implication in detail after the conclusion of the proof.

**Proposition 4.6.** *Let  $G$  be a noncompact simple Lie group. Let  $H_0 \in \bar{\mathfrak{a}}_+$  be non-zero. Let  $\Omega \subset \mathfrak{a}^* \setminus \mathfrak{a}_{\text{bad}}^*$  be compact and  $W$ -invariant. There is  $\epsilon_0 > 0$  (present in the definition of  $B^M$ ) as well as constants  $c, \tau_0 > 0$ , such that*

$$\frac{1}{\tau} \int_\tau^{2\tau} \left| \sum_{w \in W/W_M} c^M(w\lambda) \widehat{\mathbb{1}_{B^M}}(w\lambda) e^{-itw\lambda(H_0)} \right|^2 dt > c$$

for all  $\lambda \in \Omega, \tau \geq \tau_0$ .

*Proof.* We expand the square and swap the order of summation and integral to get the sum of the diagonal term

$$D(\lambda) := \sum_{w \in W/W_M} \left| c^M(w\lambda) \widehat{\mathbb{1}_{B^M}}(w\lambda) \right|^2,$$

which is independent of  $\tau$  and  $t_0$ , plus the off-diagonal contribution

$$E(\tau, \lambda) := \sum_{\substack{w, w' \in W/W_M \\ w \neq w'}} c^M(w\lambda) \overline{c^M(w'\lambda)} \widehat{\mathbf{1}_{B^M}}(w\lambda) \overline{\widehat{\mathbf{1}_{B^M}}(w'\lambda)} \frac{1}{\tau} \int_{\tau}^{2\tau} e^{-it(w\lambda - w'\lambda)(H_0)} dt.$$

We claim that for any compact  $\Omega \subset \mathfrak{a}^* \setminus \mathfrak{a}_{\text{bad}}^*$  we have

- (1)  $D(\lambda) \gg 1$  for all  $\lambda \in \Omega$ , and
- (2)  $E(\tau, \lambda) \ll \tau^{-1}$  for all  $\lambda \in \Omega$ ,

with the implied constants depending on  $\Omega$ . Taking  $\tau_0$  sufficiently large will guarantee that  $E$  is negligible with respect to  $D$  for  $\tau \geq \tau_0$ .

We begin by proving the stated lower bound on  $D(\lambda)$ . It will be sufficient to show that there is a choice of  $\epsilon_0 > 0$  such that  $\widehat{\mathbf{1}_{B^M}}(\lambda) c^M(\lambda)$  is bounded away from zero. Recall that  $c^M(\lambda)$  is holomorphic and non-vanishing on the open subset containing  $\mathfrak{a}_{\text{reg}}^*$  described in (22). Since  $\Omega \subset \mathfrak{a}_{\text{reg}}^*$  and  $\Omega$  is compact, we deduce that  $c^M(\lambda)$  is uniformly bounded away from zero on  $\Omega$ . Furthermore, we have

$$|\widehat{\mathbf{1}_{B^M}}(\lambda)| \geq |\operatorname{Re} \widehat{\mathbf{1}_{B^M}}(\lambda)| = \left| \int_{B^M} \operatorname{Re} \varphi_{-\lambda}^M(g) dg \right|.$$

As in the proof of Lemma 4.2 there exists a neighborhood  $U$  of the identity in  $G$  such that  $\operatorname{Re} \varphi_{-\lambda}(g) > 1/2$  for all  $g \in U$  and  $\lambda \in \Omega$ . We can therefore choose  $\epsilon_0 > 0$  small enough in the definition of  $B^M = K_M \exp(B_{\mathfrak{a}}(0, \epsilon_0)) K_M$  so that, for all  $\lambda \in \Omega$ ,

$$\int_{B^M} \operatorname{Re} \varphi_{-\lambda}^M(g) dg > \frac{1}{2} \operatorname{vol}(B^M).$$

We deduce that  $\widehat{\mathbf{1}_{B^M}}(\lambda)$  is bounded away from zero for all  $\lambda \in \Omega$ , proving point (1).

We now address the off-diagonal term. Firstly, since  $\lambda$  is confined to the compact  $\Omega$ , and  $\widehat{\mathbf{1}_{B^M}}(w\lambda)$  is continuous, we have that  $\widehat{\mathbf{1}_{B^M}}(w\lambda)$  is uniformly bounded on  $\Omega$ . On the other hand  $c^M(w\lambda)$  is singular on a subset of  $\mathfrak{a}_{\text{sing}}^*$  so we shall require that  $\Omega$  avoid  $\mathfrak{a}_{\text{sing}}^*$ . Therefore for  $\lambda \in \Omega$  we also have a uniform bound on  $c^M(w\lambda)$ .

We thus have

$$E(\tau, \lambda) \ll \max_{\substack{w, w' \in W/W_M \\ w \neq w'}} \frac{1}{\tau} \left| \int_{\tau}^{2\tau} e^{-it\lambda(wH_0 - w'H_0)} dt \right|.$$

Notice that  $\mathfrak{a}_{\text{bad}}^*$  is precisely the locus where the exponential phase vanishes. Since  $\lambda$  is confined to a compact  $\Omega \subset \mathfrak{a}^* \setminus \mathfrak{a}_{\text{bad}}^*$ , the phases appearing in the  $t$ -integral are all uniformly bounded away from zero. Thus, if  $X$  is any one of the differences  $wH_0 - w'H_0$ , then

$$\left| \int_{\tau}^{2\tau} e^{-it\lambda(X)} dt \right| = |\lambda(X)|^{-1} \left| 1 - e^{-i\tau\lambda(X)} \right| \leq 2|\lambda(X)|^{-1} \ll 1,$$

which establishes point (2), and completes the proof of Proposition 4.6.  $\square$

We now return to the proof of Proposition 4.3. Let  $\epsilon_0, c > 0$  be given by Proposition 4.6. The proof of Proposition 4.6 shows that we may shrink  $\epsilon_0 > 0$  while maintaining the value of  $c > 0$ . We may therefore take  $\epsilon_0 > 0$  small enough so that the  $O(\epsilon_0)$ -term in Lemma 4.4 is bounded by  $c/4$ . We are furthermore free to increase  $\tau_0$  in Proposition 4.3 as we wish. We therefore choose  $\tau_0 > 0$  large enough so that for any  $t \geq \tau_0$  the  $O(e^{-\varepsilon t})$ -term in Lemma 4.4 is

smaller than  $c/4$ . Taking  $t_0 = \tau_0$  and the above choice of  $\epsilon_0$  in the statement of Proposition 4.3 will yield the lower bound of  $c/2$  in that statement.

**4.5. Identifying  $\mathfrak{a}_{\text{bad}}^*$  for extremal  $H_0$  in classical groups.** In this section, we assume that  $G$  is simple. We wish to describe the nature of  $\mathfrak{a}_{\text{bad}}^*$  in case the directing element  $H_0 \in \bar{\mathfrak{a}}_+$  is extremal. Thus far we have only defined extremal for the classical types as well as  $E_7$  as these are the only roots systems admitting semi-dense root subsystems. However, the spectral estimate (Theorem 4.1) holds in any type. In order to have a complete analysis for all types in the following proposition, we shall define extremal elements in case the reduced root system is of type  $E_6, E_8, F_4$  or  $G_2$  via the table below. However, in subsequent sections, the term extremal will again be reserved only for those elements specified in § 3.6.

type of $\Phi_{\text{red}}$	extremal nodes $\bullet$	type of $\Phi_{M,\text{red}}$
$E_6$		$D_5$
$E_8$		$E_7$
$F_4$		$B_3$ or $C_3$
$G_2$		$A_1$

TABLE 2. Extremal nodes for the remaining exceptional types

We will show that the inclusion (35) is nearly an equality in most cases. The basic principle is that the Weyl group  $W_M$  of an extremal  $H_0$  is large, which will tend to minimize the number of distinct lines  $\mathbb{R}.(wH_0 - w'H_0) \subset \mathfrak{a}$ .

**Proposition 4.7.** *Assume that  $G$  is simple and let  $H_0 \in \bar{\mathfrak{a}}_+$  be extremal.*

- (1) *If  $\Phi_{\text{red}}$  is of type  $A_n, B_n, C_n$ , or  $G_2$  then  $\mathfrak{a}_{\text{bad}}^* = \mathfrak{a}_{\text{sing}}^*$ .*
- (2) *If  $\Phi_{\text{red}}$  is of type  $D_n$ , then*

$$\mathfrak{a}_{\text{bad}}^* = \mathfrak{a}_{\text{sing}}^* \cup W.(\mathbb{R}.\varpi_1^\vee),$$

where  $\varpi_1^\vee$  is the unique extremal fundamental coweight.

*Proof.* We must show that

- (1) If  $G$  is of type  $A_n, B_n, C_n$  or  $G_2$  then equality holds in (35);
- (2) If  $G$  is of type  $D_n$ , then equality holds in (35) with the right-hand side replaced by

$$\left( \bigcup_{\alpha \in \Phi} \mathbb{R}\alpha^\vee \right) \cup W.(\mathbb{R}.\varpi_1^\vee)$$

where  $\varpi_1^\vee$  is the extremal fundamental coweight (a multiple of  $H_0$ ).

In all cases, by assumption  $H_0$  is a non-zero multiple of an extremal fundamental coweight, which we denote by  $\varpi_0^\vee \in \widehat{\Delta}^\vee$ . Without loss of generality we can simply take  $H_0 = \varpi_0^\vee$ .

We begin by assuming that  $\Phi_{\text{red}}$  is of type  $A_n, B_n, C_n$ , or  $G_2$ . For equality to hold in (35), we must have that for any  $w \in W/W_M$ , the difference  $\varpi_0^\vee - w\varpi_0^\vee$  must lie in the line spanned by a coroot. In fact, using the computation (37), we need only to show that every  $w \in W/W_M$  with  $w \neq 1$ , can be represented by a reflection in  $W$ . First note that the reflections  $s_\alpha$  for  $\alpha \in \Phi_{\text{red}}^+ \setminus \Phi_{\text{red},M}^+$  are pairwise distinct mod  $W_M$ , that is, the image of the

set  $\{1\} \cup \{s_\alpha \mid \alpha \in \Phi_{\text{red}}^+ \setminus \Phi_{\text{red},M}^+\}$  in  $W/W_M$  has cardinality  $|\Phi_{\text{red}}^+ \setminus \Phi_{\text{red},M}^+| + 1$ . For point (1), it will therefore suffice to check that

$$|W/W_M| = |\Phi_{\text{red}}^+ \setminus \Phi_{\text{red},M}^+| + 1$$

if  $\Phi_{\text{red}}$  is of type  $A_n$ ,  $B_n$ ,  $C_n$ , or  $G_2$ . For that we compute the following table:

type of $G$	$ W $	$ W_M $	$ W/W_M $	$ \Phi_{\text{red}} $	$ \Phi_{\text{red},M} $	$ \Phi_{\text{red}}^+ \setminus \Phi_{\text{red},M}^+ $
$A_n$	$(n+1)!$	$n!$	<b>n+1</b>	$n(n+1)$	$(n-1)n$	<b>n</b>
$B_n$	$2^n n!$	$2^{n-1}(n-1)!$	<b>2n</b>	$2n^2$	$2(n-1)^2$	<b>2n-1</b>
$C_n$	$2^n n!$	$2^{n-1}(n-1)!$	<b>2n</b>	$2n^2$	$2(n-1)^2$	<b>2n-1</b>
$D_n$	$2^{n-1} n!$	$2^{n-2}(n-1)!$	<b>2n</b>	$2n(n-1)$	$2(n-1)(n-2)$	<b>2n-2</b>
$E_6$	51840	1920	<b>27</b>	72	40	<b>16</b>
$E_7$	2903040	51840	<b>56</b>	126	72	<b>27</b>
$E_8$	696729600	2903040	<b>240</b>	240	126	<b>57</b>
$F_4$	1152	48	<b>24</b>	48	18	<b>15</b>
$G_2$	12	2	<b>6</b>	12	2	<b>5</b>

An examination of the boldfaced columns of the first, second, third, and last rows concludes the proof of point (1).

We now take  $G$  to have root system  $D_n$ . Consulting Table 1, we may then take the extremely singular element  $H_0$  to be the highest fundamental coweight  $\varpi_1^\vee$ . To prove point (2), we must therefore show that, for any  $w \in W/W_M$ , the difference  $\varpi_1^\vee - w\varpi_1^\vee$  is either in a line spanned by a coroot or by  $\varpi_1^\vee = H_0$  itself. The  $2n-1$  elements  $w \in \{1\} \cup \{s_\alpha : \alpha \in \Phi_{\text{red}}^+ \setminus \Phi_{\text{red},M}^+\}$  have distinct images in the size  $2n$  quotient  $W/W_M$ , and by (37), satisfy  $\varpi_1^\vee - w\varpi_1^\vee \in \bigcup_{\alpha \in \Phi} \mathbb{R}\alpha^\vee$ . The remaining coset in  $W/W_M$ , in the notation of paragraphs (IX) and (X) of Ch. VI §4.8 of [Bou02], is represented by, say,  $s_{12}$ , which changes the sign of the first and second basis vectors  $\varepsilon_1, \varepsilon_2$ . Since  $\varepsilon_1 = \varpi_1^\vee$ , this yields  $\varpi_1^\vee - s_{12}(\varpi_1^\vee) = \varpi_1^\vee - (-\varpi_1^\vee) = 2\varpi_1^\vee$ , as required. This concludes the proof of point (2), and hence the proposition.  $\square$

## 5. GEOMETRIC ESTIMATE: STATEMENT AND REDUCTION STEPS

We deduce from Theorem 4.1 that, under the stated conditions on  $\Omega, \tau$ , and  $a$ , that

$$\sum_{j: \lambda_j \in \Omega} |\langle a\psi_j, \psi_j \rangle|^2 \ll \sum_{j,k} |\langle \mathbf{A}(\tau)\psi_j, \psi_k \rangle|^2 = \|\mathbf{A}(\tau)\|_{\text{HS}}^2,$$

where the spectral sum has been extended by positivity, and the right-hand side is the Hilbert–Schmidt norm of the operator  $\mathbf{A}(\tau)$ . We have thus reduced our problem to bounding the Hilbert–Schmidt norm of  $\mathbf{A}(\tau)$ .

The remainder of the paper is organized around the proof of Theorem 5.1 stated below.

**Theorem 5.1** (Geometric estimate). *Let  $G$  be a product of non-compact simple real Lie groups. Let  $X = G/K$ , where  $K$  is a maximal compact subgroup, be the associated symmetric space. Let  $\Gamma < G$  be a torsion free, cocompact, irreducible lattice and set  $Y = \Gamma \backslash X$ . Assume that  $G$  admits a simple factor  $G_1$  satisfying condition (1) of Theorem 1.1. Let  $H_0 \in \bar{\mathfrak{a}}_{1,+}$  be an extremal coweight for  $G_1$ , identified with an element in  $\bar{\mathfrak{a}}_+$  through  $\mathfrak{a}_1 \subset \mathfrak{a}$ .*

There are constants  $c_1, c_2, c_3, \epsilon_0 > 0$ , depending only on  $G$ , an integer  $k \in \mathbb{Z}_{\geq 0}$ , depending only on the reduced root system of  $G_1$ , and  $\theta > 0$ , depending on the integrability exponent of  $G_1$  acting on  $L_0^2(\Gamma \backslash G)$ , such that, for all  $\tau \gg 1$  and all mean-zero functions  $a \in L^\infty(Y)$ ,

$$(38) \quad \|\mathbf{A}(\tau)\|_{\text{HS}}^2 \ll \|a\|_2^2 \frac{(\log \tau)^k}{\tau \theta^2} + \frac{e^{c_1 \tau}}{\text{InjRad}(Y)^{\dim Y}} \text{vol}(Y_{\leq c_2 \tau + c_3}) \|a\|_\infty^2.$$

Here,  $\mathbf{A}(\tau)$  is defined relative to the parameters  $\epsilon_0 > 0$  and  $H_0$ .

The purpose of this section is to reduce the proof of Theorem 5.1 to Theorem 2.2 from the introduction, applied to the intersection of the spherical shell  $S_t \subset G_1$  and its translates in the irreducible symmetric space  $X_1 = G_1/K_1$ . We carry out this reduction by a thick-thin decomposition of the kernel of the operator  $\mathbf{A}(\tau)$  in Lemma 5.2, a calculation of the support of this kernel in Corollary 5.5, and, most notably, an invocation of the Nevo ergodic theorem (Proposition 5.6 below). The remaining sections of the paper will then be dedicated to the proof of Theorem 2.2.

Together, Proposition 3.4, Theorem 4.1, and Theorem 5.1 yield the estimate (1), completing the proof of Theorem 1.1 as we now explain. Let  $\mathbf{A}_n(\tau)$  denote the operator defined in (29) with respect to the test function  $a_n$ . Inserting the aforementioned results, along with  $\|a_n\|_2^2 \ll \text{vol}(\Gamma_n \backslash G) \|a_n\|_\infty^2$ ,  $\|a_n\|_\infty = O(1)$ , the uniformly discrete hypothesis on  $\Gamma_n$ , and the uniform spectral gap for  $G_1$ , we find that for all  $n$  and  $\tau$  sufficiently large

$$(39) \quad \begin{aligned} \frac{1}{N(\Gamma_n, \Omega)} \sum_{\lambda_j^{(n)} \in \Omega} |\langle a_n \psi_j^{(n)}, \psi_j^{(n)} \rangle|^2 &\ll \frac{1}{\text{vol}(\Gamma_n \backslash G)} \sum_{\lambda_j^{(n)} \in \Omega} |\langle a_n \psi_j^{(n)}, \psi_j^{(n)} \rangle|^2 \\ &\ll \frac{1}{\text{vol}(\Gamma_n \backslash G)} \|\mathbf{A}_n(\tau)\|_{\text{HS}}^2 \\ &\ll \frac{(\log \tau)^k}{\tau} + e^{c_1 \tau} \frac{\text{vol}((Y_n)_{\leq c_2 \tau + c_3})}{\text{vol}(Y_n)}. \end{aligned}$$

We can choose for each  $n$  a choice of  $\tau_n$  such that the above expression goes to zero. So long as  $\tau_n \rightarrow \infty$  the first term will go to zero. It follows from the Benjamini–Schramm convergence of  $Y_n$  towards  $X$  that there exists a sequence of  $R_n$  tending toward infinity such that

$$\alpha_n := \frac{\text{vol}((Y_n)_{\leq R_n})}{\text{vol}(Y_n)} \rightarrow 0.$$

Let  $r_n$  be a sequence tending to infinity such that  $r_n + c_3 \leq R_n$  and  $e^{c_1 r_n / c_2} \alpha_n \rightarrow 0$ . Then, setting  $\tau_n = r_n / c_2$ , the second term in (39) is

$$e^{c_1 r_n / c_2} \frac{\text{vol}((Y_n)_{\leq r_n + c_3})}{\text{vol}(Y_n)} \leq e^{c_1 r_n / c_2} \alpha_n \rightarrow 0,$$

as desired.

**5.1. A general bound on Hilbert–Schmidt norms.** We begin with a very general upper bound on the Hilbert–Schmidt norm of an integral operator on  $L^2(\Gamma \backslash G)$ .

**Lemma 5.2.** *Let  $G$  be a product of non-compact simple real Lie groups. Let  $A$  be a measurable function on  $X \times X$  which is invariant under the left diagonal action. For a uniform*

lattice  $\Gamma < G$  let  $\mathbf{A}$  be the integral operator on  $L^2(\Gamma \backslash G)$  with kernel  $\sum_{\gamma \in \Gamma} A(g, \gamma h)$ . Then there is  $c > 0$ , depending only on  $G$ , such that

$$\|\mathbf{A}\|_{\text{HS}}^2 \leq \int_{\Gamma \backslash G} \int_G |A(g, h)|^2 dg dh + \frac{e^{cT}}{\text{InjRad}(Y)^{\dim Y}} \text{vol}(Y_{\leq 2T}) \|A\|_{\infty}^2,$$

for all  $T > 0$  satisfying  $\{g^{-1}h : (g, h) \in \text{supp}(A)\} \subset K \exp(B_{\mathfrak{a}}(0, T))K$ .

*Proof.* See Lemma 5.2 of [BM23], which is an adaptation of Lemma 5.1 of [LMS17] to the general setting.  $\square$

To prove Theorem 5.1, we shall apply Lemma 5.2 with the integral operator  $\mathbf{A}(\tau)$ , defined in (29). This will require examining its kernel, which is described in the following result.

Throughout this subsection and the next, we may work under more general hypotheses on  $G$  and  $H_0$  than those imposed by the conditions of Theorem 5.1. The condition on the subsystem of reduced roots of  $G_1$  and the extremal coweight hypothesis on  $H_0$  will only play a role starting in subsection 5.3.

**Lemma 5.3.** *Let  $A(\tau)$  be the function on  $\Gamma \backslash (G \times G)$  given by*

$$A(\tau)(g, h) = \frac{1}{\tau} \int_{\tau}^{2\tau} e^{-2t\rho(H_0)} \int_{gE_t \cap hE_t} a(x) dx dt.$$

*Then the kernel of  $\mathbf{A}(\tau)$ , as an integral operator on  $L^2(\Gamma \backslash G)$ , is given by  $\sum_{\gamma \in \Gamma} A(\tau)(g, \gamma h)$ .*

*Proof.* By definition, for a function  $f \in L^2(\Gamma \backslash G)$ ,

$$(U_t a U_t^*) f(g) = e^{-2t\rho(H_0)} \int_{E_t} a(gh_1) \int_{E_t^{-1}} f(gh_1 h_2) dh_2 dh_1.$$

We change variables in the inner sum by setting  $h = gh_1 h_2$  and use Fubini's theorem to get

$$\begin{aligned} (U_t a U_t^*) f(g) &= e^{-2t\rho(H_0)} \int_{E_t} a(gh_1) \int_G f(h) \mathbb{1}_{E_t^{-1}}(h_1^{-1} g^{-1} h) dh dh_1 \\ &= e^{-2t\rho(H_0)} \int_G f(h) \int_G a(gh_1) \mathbb{1}_{E_t^{-1}}(h_1^{-1} g^{-1} h) \mathbb{1}_{E_t}(h_1) dh_1 dh. \end{aligned}$$

We then change variables via  $x = gh_1$  to get

$$\begin{aligned} (U_t a U_t^*) f(g) &= e^{-2t\rho(H_0)} \int_G f(h) \int_G a(x) \mathbb{1}_{E_t^{-1}}(x^{-1} h) \mathbb{1}_{E_t}(g^{-1} x) dx dh \\ &= e^{-2t\rho(H_0)} \int_G f(h) \int_{gE_t \cap hE_t} a(x) dx dh \\ &= e^{-2t\rho(H_0)} \int_{\Gamma \backslash G} f(h) \sum_{\gamma \in \Gamma} \int_{gE_t \cap \gamma h E_t} a(x) dx dh. \end{aligned}$$

Averaging over  $t$  yields the result.  $\square$

**5.2. The support of the kernel function.** To obtain the second term in the estimate (38) via an application of Lemma 5.2, it will be necessary to calculate the real parameter  $T$  which bounds the size of the support of the kernel.

It follows from Lemma 5.3, and the bi- $K$ -invariance of the sets  $E_t$ , that

$$(40) \quad \{g^{-1}h : (g, h) \in \text{supp}(A(\tau))\} \subseteq \bigcup_{\tau \leq t \leq 2\tau} \{g^{-1}h : gE_t \cap hE_t \neq \emptyset\} = \bigcup_{\tau \leq t \leq 2\tau} K\{e^H : e^H E_t \cap E_t \neq \emptyset\} K.$$

We therefore need to estimate the Euclidean norm of the largest element  $H \in \bar{\mathfrak{a}}_+$  for which  $e^H E_t$  intersects  $E_t$ . Now, recall from (28) that  $E_t$  factorizes as  $S_t \times B \subset G_1 \times \mathcal{G}_2$ , so that

$$(41) \quad e^H E_t \cap E_t = (e^{H_1} S_t \cap S_t)(e^{H_2} B \cap B), \quad \text{where } H = H_1 + H_2 \in \mathfrak{a}_1 \oplus \text{Lie}(\mathcal{A}_2).$$

We may take  $\epsilon_0 > 0$  small enough in the definition (27) of  $B$  so that  $e^{H_2} B \cap B = \emptyset$  if  $\|H_2\| \geq 1$ . It therefore suffices to estimate the Euclidean norm of the largest element  $H \in \bar{\mathfrak{a}}_{1,+}$  for which  $e^H S_t$  intersects  $S_t$ . For this reason, for the rest of this subsection, we may and shall assume that  $G = G_1$  is a non-compact simple Lie group and  $E_t = S_t$ .

We begin by proving the following triangle inequality in the irreducible symmetric space  $X$ . Given  $H \in \mathfrak{a}$ , let  $\text{Conv}(W.H)$  denote the convex hull of the  $W$ -orbit of  $H$ . We define a partial order on  $\bar{\mathfrak{a}}_+$  via  $H_1 \preceq H_2$  if and only if  $\text{Conv}(W.H_1) \subseteq \text{Conv}(W.H_2)$ . We define a “Weyl chamber-valued metric” on  $G/K$  via  $d_{\bar{\mathfrak{a}}_+}(gK, hK) = H \in \bar{\mathfrak{a}}_+$  if  $g^{-1}h \in K e^H K$ .

**Lemma 5.4.** *For all  $x, y, z \in G/K$ , we have*

$$d_{\bar{\mathfrak{a}}_+}(x, z) \preceq d_{\bar{\mathfrak{a}}_+}(x, y) + d_{\bar{\mathfrak{a}}_+}(y, z).$$

*Proof.* We first reduce the proof to the corresponding Lie algebra version, using [KLM08]. Let  $\mathfrak{p}$  be the orthogonal complement to  $\mathfrak{k}$  with respect to the Cartan involution; thus  $\mathfrak{p}$  is invariant under  $\text{Ad}(K)$ . We have a Cartan decomposition with respect to adjoint  $K$ -action on  $\mathfrak{p}$ , namely  $\mathfrak{p} = \bigsqcup_{H \in \bar{\mathfrak{a}}_+} K.H$ .

Let  $\kappa : \mathfrak{p} \rightarrow \bar{\mathfrak{a}}_+$  denote the corresponding Cartan projection, and let  $d_{\bar{\mathfrak{a}}_+}^{\text{inf}} : \mathfrak{p} \times \mathfrak{p} \rightarrow \bar{\mathfrak{a}}_+$ ,  $d_{\bar{\mathfrak{a}}_+}^{\text{inf}}(X, Y) = \kappa(Y - X)$ , denote the associated Weyl chamber-valued metric; this metric is invariant under the action of the Cartan motion group  $K \ltimes \mathfrak{p}$ . According to Main Theorem 1.2 of [KLM08], given three points  $x, y, z \in G/K$  such that  $d_{\bar{\mathfrak{a}}_+}(x, y) = \lambda_1$ ,  $d_{\bar{\mathfrak{a}}_+}(y, z) = \lambda_2$  and  $d_{\bar{\mathfrak{a}}_+}(x, z) = \lambda_3$ , there exist three points  $X, Y, Z \in \mathfrak{p}$  such that  $d_{\bar{\mathfrak{a}}_+}^{\text{inf}}(X, Y) = \lambda_1$ ,  $d_{\bar{\mathfrak{a}}_+}^{\text{inf}}(Y, Z) = \lambda_2$  and  $d_{\bar{\mathfrak{a}}_+}^{\text{inf}}(X, Z) = \lambda_3$  (and vice versa). Thus, it suffices to show that for any  $X, Y, Z \in \mathfrak{p}$  we have that

$$d_{\bar{\mathfrak{a}}_+}^{\text{inf}}(X, Z) \preceq d_{\bar{\mathfrak{a}}_+}^{\text{inf}}(X, Y) + d_{\bar{\mathfrak{a}}_+}^{\text{inf}}(Y, Z),$$

which is in turn equivalent to showing that for any  $X, Y \in \mathfrak{p}$  we have that

$$(42) \quad \kappa(X + Y) \preceq \kappa(X) + \kappa(Y).$$

Recall that the Killing form on  $\mathfrak{p}$  is positive definite. Let  $\Pi_{\mathfrak{a}} : \mathfrak{p} \rightarrow \mathfrak{a}$  denote orthogonal projection with respect to the Killing form. By the Kostant convexity theorem [Kos73] (Theorem 8.2), for any  $X \in \mathfrak{p}$ , we have  $\Pi_{\mathfrak{a}}(\text{Ad}(K).X) = \text{Conv}(W.X)$ . Now, given  $X, Y \in \mathfrak{p}$ , we wish to show (42). By applying an element of  $K$ , we may assume that  $X + Y \in \bar{\mathfrak{a}}_+$ , and thus  $\kappa(X + Y) = X + Y$ . We also have that  $X = k_1 \cdot \kappa(X)$  and  $Y = k_2 \cdot \kappa(Y)$  for some  $k_i \in K$ .

We thus have

$$\begin{aligned}
\kappa(X + Y) &= X + Y \\
&= \Pi_{\mathfrak{a}}(X + Y) \\
&= \Pi_{\mathfrak{a}}(X) + \Pi_{\mathfrak{a}}(Y) \\
&= \Pi_{\mathfrak{a}}(k_1 \cdot \kappa(X)) + \Pi_{\mathfrak{a}}(k_2 \cdot \kappa(Y)) \\
&\in \text{Conv}(W \cdot \kappa(X)) + \text{Conv}(W \cdot \kappa(Y)) \\
&\subseteq \text{Conv}(W \cdot (\kappa(X) + \kappa(Y))),
\end{aligned}$$

as required.  $\square$

Let  $H_0 \in \bar{\mathfrak{a}}_+$  be non-zero. Let

$$(43) \quad P = \text{Conv}(W \cdot (H_0 - w_0 H_0)) \cap \bar{\mathfrak{a}}_+,$$

where  $w_0$  is the longest word of the Weyl group. For  $t > 0$  write  $P_t = tP$ . We define a “polytopal norm” on  $\bar{\mathfrak{a}}_+$  by putting  $\|H\|_P := \inf\{t : H \in P_t\}$ .

**Corollary 5.5.** *There is  $\epsilon_0 > 0$  small enough (in the definition of  $S_t$ ) such that any  $H \in \bar{\mathfrak{a}}_+$  for which  $e^H S_t \cap S_t \neq \emptyset$  satisfies  $\|H\|_P \leq t + 1$ . In particular,*

$$\{g^{-1}h : (g, h) \in \text{supp}(A(\tau))\} \subset K \exp(B_{\mathfrak{a}}(0, T))K,$$

where  $T = (\tau + 1)\|H_0 - w_0 H_0\|$ .

*Proof.* We may take  $\epsilon_0$  small so that  $B_{\mathfrak{a}}(0, \epsilon_0) \subset \text{Conv}(W \cdot H_0)$ , and thus  $B_{\mathfrak{a}}(tH_0, \epsilon_0) \subset \text{Conv}(W \cdot (t + 1)H_0)$ . We apply Lemma 5.4 with  $y \in e^H S_t \cap S_t$ ,  $x = K$ , and  $z = e^H K$ , to obtain

$$H = d_{\bar{\mathfrak{a}}_+}(K, e^H K) \preceq d_{\bar{\mathfrak{a}}_+}(K, y) + d_{\bar{\mathfrak{a}}_+}(y, e^H K) \preceq (t + 1)H_0 + (t + 1)(-w_0 H_0).$$

It follows that  $H \in P_{t+1}$ . To obtain the second statement, we apply (40) and the fact that  $P_t$  is contained in the Euclidean ball in  $\mathfrak{a}$  of radius  $t\|H_0 - w_0 H_0\|$ .  $\square$

Let  $c > 0$  be as in Lemma 5.2. If we set  $c_1 = c\|H_0 - w_0 H_0\|$  and  $c_2 = c_3 = 2\|H_0 - w_0 H_0\|$ , we obtain the second term in (38).

**5.3. Main term bound.** We denote the main term in Lemma 5.2 by

$$(44) \quad M(\tau) = \int_{\Gamma \backslash G} \int_G |A(\tau)(g, h)|^2 dg dh.$$

The goal of this subsection is to show that, under the assumptions laid out in Theorem 5.1, and assuming Theorem 2.2, that  $M(\tau)$  is bounded by the first term in the estimate (38). This will then fully reduce the proof of Theorem 5.1 to that of Theorem 2.2.

We shall make use of the following ergodic theorem due to Nevo [Nev98]:

**Proposition 5.6** (Nevo). *Let  $G$  be a non-compact simple real Lie group with finite center acting by measure-preserving transformations on a probability space  $(\Sigma, \mu)$ . Let  $r_\Sigma$  denote the action on  $L^2(\Sigma, \mu)$  and  $r_\Sigma^0$  the restriction to the orthocomplement  $L_0^2(\Sigma, \mu)$  to the constant functions. Assume that  $r_\Sigma^0$  has a spectral gap. There exist constants  $\theta, C > 0$ , depending on*

the integrability exponent for  $r_\Sigma^0$ , such that for any measurable set  $E \subset G$  of finite measure, we have

$$\|r_\Sigma^0(\text{vol}(E)^{-1}\mathbf{1}_E)\|_{\text{op}} \leq C\text{vol}(E)^{-\theta},$$

where  $\|\cdot\|_{\text{op}}$  denotes the operator norm.

*Remark 5.7.* In Section 4.1 of [GN15] a sketch of the proof of Proposition 5.6 is given. From that discussion one sees that the best possible admissible value of  $\theta$  by the proof technique presented there is  $\theta = 1/4$ , corresponding to  $r_\Sigma^0$  being tempered.

The following preliminary bound is valid under more general assumptions on  $G$  and  $H_0$  than those laid out in Theorem 5.1.

**Proposition 5.8.** *Let  $G$  be a product of non-compact simple real Lie groups and  $\Gamma < G$  an irreducible lattice. Fix a simple factor  $G_1$  of  $G$  and let  $H_0 \in \overline{\mathfrak{a}_{1+}}$  be non-zero.*

*There is  $\theta > 0$ , depending on the integrability exponent of  $G_1$  acting on  $L_0^2(\Gamma \backslash G)$ , and  $\epsilon_0 > 0$  (implicit in the definition of  $S_t$  and  $E_t$ ), such that for  $\tau > 1$ , and mean-zero  $a \in L^\infty(Y)$ ,*

$$M(\tau) \ll \frac{\|a\|_2^2}{\tau^2} \int_{P_{\tau+1}} e^{2\rho(H)} \left| \int_{\max\{\tau, \|H\|_P - 1\}}^{2\tau} e^{-2t\rho(H_0)} \text{vol}(e^H S_t \cap S_t)^{1-\theta} dt \right|^2 dH,$$

where the  $H$ -integral is taken over  $P_{\tau+1} \subset G_1$ .

*Proof.* For a measurable set  $E \subset G$  it will be convenient to let  $\varrho_{\Gamma \backslash G}(E) = \varrho_{\Gamma \backslash G}(\text{vol}(E)^{-1}\mathbf{1}_E)$ . Inserting Lemma 5.3 into the definition of  $M(\tau)$  and changing variables  $x \mapsto hx$  and  $g \mapsto h^{-1}g$ , we have

$$\begin{aligned} M(\tau) &= \frac{1}{\tau^2} \int_{\Gamma \backslash G} \int_G \left| \int_{\tau}^{2\tau} e^{-2t\rho(H_0)} \int_{gE_t \cap hE_t} a(x) dx dt \right|^2 dg dh \\ &= \frac{1}{\tau^2} \int_{\Gamma \backslash G} \int_G \left| \int_{\tau}^{2\tau} e^{-2t\rho(H_0)} \int_{gE_t \cap E_t} a(hx) dx dt \right|^2 dg dh. \end{aligned}$$

Using Cartan decomposition for the integral over  $g \in G$  and the fact that  $kE_t = E_t$  for all  $k \in K$ , this equals

$$\frac{1}{\tau^2} \int_{\Gamma \backslash G} \int_K \int_{\mathfrak{a}_+} \left| \int_{\tau}^{2\tau} e^{-2t\rho(H_0)} \int_{ke^H E_t \cap E_t} a(hx) dx dt \right|^2 J(H) dH dk dh.$$

Changing  $x \mapsto kx$  and using again that  $k^{-1}E_t = E_t$ , we get

$$\frac{1}{\tau^2} \int_{\Gamma \backslash G} \int_K \int_{\mathfrak{a}_+} \left| \int_{\tau}^{2\tau} e^{-2t\rho(H_0)} \int_{e^H E_t \cap E_t} a(hkx) dx dt \right|^2 J(H) dH dk dh.$$

Changing variables  $hk \mapsto h$ , we can absorb the integral over  $k \in K$  into the integral over  $h \in \Gamma \backslash G$ . Further recalling that  $J(H) \ll e^{2\rho(H)}$ , the above expression is then bounded by

$$\frac{1}{\tau^2} \int_{\Gamma \backslash G} \int_{\mathfrak{a}_+} e^{2\rho(H)} \left| \int_{\tau}^{2\tau} e^{-2t\rho(H_0)} \text{vol}(e^H E_t \cap E_t) (\varrho_{\Gamma \backslash G}(e^H E_t \cap E_t) a)(h) dt \right|^2 dH dh.$$

We write  $H = H_1 + H_2 \in \mathfrak{a}_1 \oplus \text{Lie}(\mathcal{A}_2)$  as in (41). Taking  $\epsilon_0$  small enough in the definition of  $B \subset \mathcal{G}_2$ , we may truncate the  $H_2$  integral to  $\|H_2\| \leq 1$  and apply  $\text{vol}(e^{H_2} B \cap B) \ll 1$ . From Corollary 5.5 we may truncate the  $H_1$  integral to  $\|H_1\|_P \leq \tau + 1$ . By the same token,

given  $H_1 \in P_{\tau+1}$ , the  $t$  for which  $e^{H_1} S_t \cap S_t \neq \emptyset$  satisfy  $\|H_1\|_P - 1 \leq t$ . Inserting this, and applying the Minkowski integral inequality, we obtain an upper bound of the form

$$\frac{1}{\tau^2} \int_{P_{\tau+1} \times \mathcal{B}_2(0,1)} e^{2\rho(H_1)} \left| \int_{\max\{\tau, \|H_1\|_P - 1\}}^{2\tau} e^{-2t\rho(H_0)} \text{vol}(e^{H_1} S_t \cap S_t) \|\varrho_{\Gamma \setminus G}(e^H E_t \cap E_t)(a)\|_2 dt \right|^2 dH.$$

Let  $\varrho_{\Gamma \setminus G}^0$  denote the right-regular representation of  $G$  on  $L_0^2(\Gamma \setminus G)$ . Since  $a \in L_0^2(\Gamma \setminus G)$ , we have

$$\|\varrho_{\Gamma \setminus G}(e^H E_t \cap E_t)a\|_2 = \|\varrho_{\Gamma \setminus G}^0(e^H E_t \cap E_t)a\|_2 \leq \|a\|_2 \|\varrho_{\Gamma \setminus G}^0(e^H E_t \cap E_t)\|_{\text{op}},$$

where  $\varrho_{\Gamma \setminus G}^0(E)$  is defined similarly as  $\varrho_{\Gamma \setminus G}(E)$ . From the spectral decomposition (25), we have

$$\|\varrho_{\Gamma \setminus G}^0(e^H E_t \cap E_t)\|_{\text{op}} = \sup_{\pi \subset L_0^2(\Gamma \setminus G)} \|\pi(e^H E_t \cap E_t)\|_{\text{op}}.$$

Since  $\pi$  is irreducible, it is a tensor product of irreducible representations  $\pi_i$  of the simple factors  $G_1, \dots, G_s$  of  $G$ . We write  $\pi = \pi_1 \otimes \sigma$ , where  $\sigma = \pi_2 \otimes \dots \otimes \pi_s$  is an irreducible representation of  $\mathcal{G}_2$ . Then, from (41),  $\|\pi(e^H E_t \cap E_t)\|_{\text{op}}$  is

$$\|\pi_1(e^{H_1} S_t \cap S_t)\|_{\text{op}} \|\sigma(e^{H_2} B \cap B)\|_{\text{op}} \ll \|\pi_1(e^{H_1} S_t \cap S_t)\|_{\text{op}} \leq \|\varrho_{\Gamma \setminus G}^0|_{G_1}(e^{H_1} S_t \cap S_t)\|_{\text{op}}.$$

Here we have used the trivial bound  $\|\sigma(e^{H_2} B \cap B)\|_{\text{op}} \leq \|\mathbf{1}_{e^{H_2} B \cap B}\|_{L^1(G)}$  (independently of  $\sigma$ ).

Since  $\Gamma$  is irreducible in  $G$ , the restriction of  $\varrho_{\Gamma \setminus G}^0$  to any simple factor has a spectral gap; see [KS09, p.3] for details. Proposition 5.6, applied to the simple factor  $G_1$  acting by right-translation on  $\Sigma = \Gamma \setminus G$ , implies that there is  $\theta > 0$ , depending on the integrability exponent of  $G_1$  acting on  $L_0^2(\Gamma \setminus G)$ , such that

$$\|\varrho_{\Gamma \setminus G}^0|_{G_1}(e^{H_1} S_t \cap S_t)\|_{\text{op}} \ll \text{vol}(e^{H_1} S_t \cap S_t)^{-\theta}.$$

Inserting this (and renaming  $H_1$  to  $H$ ) yields the statement.  $\square$

We now make full avail of the assumptions on  $G$  and  $H_0$  in Theorem 5.1 to conclude the reduction of its proof to Theorem 2.2.

**Proposition 5.9.** *Let  $G$  and  $H_0$ , and the simple factor  $G_1$ , be as in the statement of Theorem 5.1. There is a positive integer  $k$ , depending only on the reduced root system of  $G_1$ , and a constant  $\theta > 0$ , depending only on the integrability exponent of  $G_1$  acting on  $L_0^2(\Gamma \setminus G)$ , such that for all  $\tau \gg 1$ ,*

$$M(\tau) \ll \frac{(\log \tau)^k}{\tau \theta^2} \|a\|_2^2.$$

*Proof.* If  $M$  is the centralizer of  $H_0$  in  $G_1$ , it follows from Proposition 2.1(2) that  $\Phi_{\text{red}, M}$  is a semi-dense root system in the reduced roots of  $G_1$ . Taking  $\tau$  sufficiently large, we may insert Theorem 2.2 into Proposition 5.8, to get

$$M(\tau) \ll \frac{\|a\|_2^2}{\tau^2} \int_{P_{\tau+1}} e^{2\theta\rho(H)} \left| \int_{\max\{\tau, \|H\|_P - 1\}}^{2\tau} (\log t)^k e^{-2t\rho(H_0)} dt \right|^2 dH.$$

The square of the inner integral is bounded up to a multiplicative constant by

$$\frac{(\log \tau)^{2k}}{\theta^2} e^{-4\theta\|H\|_P\rho(H_0)},$$

so that for  $\tau \geq 1$ ,

$$(45) \quad M(\tau) \ll \frac{(\log \tau)^{2k}}{\tau^2 \theta^2} \|a\|_2^2 \int_{P_{2\tau}} e^{2\theta\rho(H-2\|H\|_P H_0)} dH.$$

We may then apply Corollary 5.11 below, and rename  $2k$  to  $k$ , to conclude the proof.  $\square$

**5.4. Application of the degenerate Brion's formula.** Let  $F_j \subset \bar{\mathfrak{a}}_+$ ,  $1 \leq j \leq J$ , denote the codimension one faces of  $P$  whose relative interior lies in  $\mathfrak{a}_+$ , and define the associated polyhedral cones  $C_j = \{\lambda H : H \in F_j, \lambda \in \mathbb{R}_+\}$ . Then  $\bar{\mathfrak{a}}_+ = \cup_j C_j$  and the integral in (45) can be broken up as

$$\int_{P_{2\tau}} e^{2\theta\rho(H-2\|H\|_P H_0)} dH = \sum_{j=1}^J I_j(\tau), \quad I_j(\tau) = \int_{P_{2\tau} \cap C_j} e^{2\theta\rho(H-2\|H\|_P H_0)} dH.$$

Because  $\|H\|_P$  restricted to  $C_j$  is simply a linear functional, the integrals  $I_j(\tau)$  could be expressed, via Brion's formula [Bri88; Bar93], as a weighted sum of the value of the integrand at the extremal vertices of  $P_{2\tau} \cap C_j$ , *provided that the values of the exponent at adjacent vertices are distinct*. This property does not, however, hold for  $I_j(\tau)$ , as we presently show.

Observe that the vertices of  $P_{2\tau} \cap C_j$  are simply the vertices of  $P \cap C_j$  dilated by a factor of  $2\tau$ . That both  $0$  and  $H_0 - w_0 H_0$  are adjacent vertices of  $P$  follows directly from the definition of  $P$  in (43) and the definition of  $C_j$  above. Now, the exponent in  $I_j(\tau)$  clearly vanishes at  $0$ . Moreover, since  $\|v\|_P = 1$  at every vertex  $v$  of  $P \cap C_j$  other than  $0$ , the exponent at such a vertex is

$$\rho(v) - 2\rho(H_0) = \rho(v) - \rho(H_0 - w_0 H_0),$$

which then clearly also vanishes at  $v = H_0 - w_0 H_0$ .

In preparation for a more general form of Brion's formula, which allows for the exponent at adjacent vertices to be the same, we now show that  $\rho(v) - \rho(H_0 - w_0 H_0) < 0$  for all other vertices of  $P \cap C_j$ .

**Lemma 5.10.** *Let  $H \in \bar{\mathfrak{a}}_+$ . Then  $H$  is the unique vertex of  $\text{Conv}(W.H)$  maximizing the pairing with  $\rho$ . In particular,  $\rho(v) - 2\rho(H_0) < 0$  for all vertices  $v$  of the polytope  $P \cap C_j$  other than  $0$  and  $H_0 - w_0 H_0$ .*

*Proof.* Denote by  $L$  the centralizer of  $H$  in  $G$  and by  $\Phi_L$  the roots of  $A$  in  $L$ . Note that, for any  $w \in W$ , we have  $\rho - w^{-1}\rho = \sum_{\alpha \in \Phi^+, w\alpha < 0} \alpha$  so that  $\rho(H) - \rho(wH) = (\rho - w^{-1}\rho)(H) = \sum_{\alpha \in \Phi^+, w\alpha < 0} \alpha(H) \geq 0$ . If equality holds then  $\{\alpha \in \Phi^+ : w\alpha < 0\} \subset \Phi_L$ , since  $\Phi_L$  consists precisely of those roots that vanish on  $H$ . But then  $w \in W_L$  by (9) of Proposition 3.1.  $\square$

For  $\lambda \in \mathfrak{a}^*$ , the degenerate Brion's formula [Pet24] implies that

$$(46) \quad \int_{P_{2\tau} \cap C_j} e^{\lambda(H)} dH = \sum_{\substack{\text{faces } F \text{ of } P \cap C_j \text{ on} \\ \text{which } \lambda(F) = \text{constant}}} c_F \text{vol}(F) (2\tau)^{\dim(F)} e^{2\tau\lambda(F)},$$

where the constants  $c_F$  are independent of  $\tau$ , and  $c_F$  is positive on the maximal dimensional face maximizing  $\lambda$ . Let  $\ell_j \in \mathfrak{a}^*$  denote the unique linear functional so that  $\ell_j|_{F_j} = 1$ . Then  $\|H\|_P = \ell_j(H)$  for all  $H \in P \cap C_j$ . We apply (46) with  $\lambda = 2\theta(\rho - 2\rho(H_0)\ell_j) \in \mathfrak{a}^*$  and use the preceding discussion to obtain the following result.

**Corollary 5.11.** *We have  $I_j(\tau) \ll \tau$ .*

This concludes the proof of Proposition 5.9 and therefore of Theorem 5.1, subject to the verification of Theorem 2.2.

## 6. INTERSECTION VOLUMES

In this section we present the proof of Theorem 2.2 which bounds the intersection volume  $\text{vol}(e^H S_t \cap S_t)$ , under suitable assumptions on  $G$  and  $H_0$ . We shall usually assume that  $G$  is semisimple, as most of the results hold in this generality, and indicate the places where we need to impose the condition of  $G$  being simple. We extend the definition of  $S_t$  from (27) to the case of  $G$  semisimple in the obvious way.

**6.1. An initial bound due to J.-P. Anker.** We first present a bound on  $\text{vol}(e^H S_t \cap S_t)$  communicated to us by J.-P. Anker which gives a weaker bound but with much less work. This will serve as a brief illustration of how harmonic analytic information, such as bounds on the spherical function and the  $c$ -function, can be used to estimate the geometric intersection volume, and will act as a yardstick to measure the extent to which a more elementary approach falls short of the bounds in Theorem 2.2. In this subsection with take  $H_0, H \in \bar{\mathfrak{a}}_+$  to be arbitrary. We first recall that, for  $\lambda \in \mathfrak{a}^*$  we have the bound

$$|\varphi_\lambda(e^H)| \leq \varphi_0(e^H).$$

See for instance Prop. 4.6.1 of [GV88].

We thus have

$$\begin{aligned} \text{vol}(e^H S_t \cap S_t) &= (\mathbf{1}_{S_t} * \mathbf{1}_{S_t}^\vee)(e^H) \\ &= \int_{\mathfrak{a}_+^*} |\widehat{\mathbf{1}}_{S_t}(\lambda)|^2 \varphi_\lambda(e^H) |c(\lambda)|^2 d\lambda \\ &\leq \varphi_0(e^H) \int_{\mathfrak{a}_+^*} |\widehat{\mathbf{1}}_{S_t}(\lambda)|^2 |c(\lambda)|^2 d\lambda \\ &= \varphi_0(e^H) \|\mathbf{1}_{S_t}\|_{L^2}^2 = \varphi_0(e^H) \text{vol}(S_t). \end{aligned}$$

We have that  $\text{vol}(S_t) \asymp e^{2\rho(tH_0)}$  and by Anker [Ank87] we have

$$\varphi_0(e^H) \asymp e^{-\rho(H)} \prod_{\alpha \in \Phi_{\text{red}}^+} (1 + \alpha(H)).$$

Letting  $L$  denote the centralizer of  $H$  in  $G$ , we thus get

$$(47) \quad \text{vol}(e^H S_t \cap S_t) \ll e^{\rho(2tH_0 - H)} \prod_{\alpha \in \Phi_{\text{red}}^+} (1 + \alpha(H)) \ll e^{\rho(2tH_0 - H)} \|H\|^{|\Phi_{\text{red}}^+ \setminus \Phi_{\text{red},L}^+|},$$

as all roots in  $\Phi_{\text{red},L}^+$  vanish on  $H$ .

Inserting (47) (applied to the simple factor  $G_1$  of  $G$ ), instead of Theorem 2.2, into the analysis leading to (45) would yield

$$M(\tau) \ll \frac{\|a\|_2^2}{\tau^2 \theta^2} \int_{P_{2\tau}} \|H\|^{2|\Phi_{\text{red}}^+ \setminus \Phi_{\text{red},L}^+| (1-\theta)} e^{\theta \rho(H - 2\|H\|_P H_0)} dH.$$

Here  $\Phi_{\text{red}}^1$  and  $\Phi_{\text{red}}^{1+}$  are the reduced roots and positive reduced roots of the simple factor  $G_1$ , and  $\Phi_{\text{red},L_1}^+$  are the positive reduced roots of  $L_1$ , the centralizer of  $H$  in  $G_1$ . Now Remark 5.7 shows that  $2|\Phi_{\text{red}}^{1+} \setminus \Phi_{\text{red},L_1}^+|(1-\theta) \geq 1$  for all  $H \neq 0$ , in which case the above integral is at least  $\tau^2$ , killing the decay in  $\tau$ . We conclude that a bound such as (47) would not suffice to complete the proof of Theorem 1.1.

The point of the remainder of this paper is to strengthen this bound to replace the polynomial factor in  $H$  with a log factor in case  $H_0$  is extremely singular.

**6.2. The endpoint Kunze–Stein phenomenon.** The Kunze–Stein phenomenon [KS60] says that if  $G$  is a semisimple real Lie group, then given  $f \in L^p(G)$  with  $p \in [1, 2)$  and  $g \in L^2(G)$ , we have

$$\|f * g\|_2 \leq c_p \|f\|_p \|g\|_2.$$

In fact the Kunze–Stein phenomenon is one of the main ingredients powering the Nevo ergodic theorem [Nev98] (i.e., Proposition 5.6). In case  $G$  has real rank one, Ionescu [Ion00] strengthened this to an endpoint estimate

$$\|f * g\|_{2,\infty} \leq c \|f\|_{2,1} \|g\|_{2,1},$$

where  $\|\cdot\|_{p,q}$  is the  $(p, q)$  Lorentz norm. However, as remarked in [Ion00], this bound fails in higher rank. It was noticed by the second author that if  $G$  is a group satisfying the endpoint Kunze–Stein phenomenon for bi- $K$ -invariant functions, then  $\text{vol}(e^H S_t \cap S_t) \ll e^{\rho(2tH_0 - H)}$  for any choice of  $H_0$  and  $H$ . A simplified argument of this implication was communicated to us by J.-P. Anker. The fact that we are able to get such a bound up to log factors in higher rank in case of  $H_0$  extremely singular suggests the possibility of a version of the endpoint Kunze–Stein phenomenon holding true in higher rank for bi- $K$ -invariant functions supported near extremely singular elements.

We now present J.-P. Anker’s argument. For the purposes of this subsection, we shall deviate slightly from our earlier notation. We define:

$$S_{H,r} := K \exp(B_{\mathfrak{a}}(H, r)) K.$$

Let  $H_0, H \in \bar{\mathfrak{a}}_+$ . We first claim that

$$e^H S_{H_0, r} \subset e^{H+Z} S_{H_0, 2r}$$

for any  $Z \in B_{\mathfrak{a}}(0, r)$ . This follows from the triangle inequality Lemma 5.4. More specifically: suppose  $x \in e^H S_{H_0, r}$ . Then

$$d_{\bar{\mathfrak{a}}_+}(e^{H+Z}, x) \preceq d_{\bar{\mathfrak{a}}_+}(e^{H+Z}, e^H) + d_{\bar{\mathfrak{a}}_+}(e^H, x).$$

The first term on the RHS lies in  $B_{\mathfrak{a}}(0, r)$ , and the second term lies in  $B_{\mathfrak{a}}(H_0, r)$ . Therefore the sum lies in  $B_{\mathfrak{a}}(H_0, 2r)$  as we wished to show.

This in turn implies that

$$(48) \quad \text{vol}(e^H S_{H_0, r} \cap S_{H_0, r}) \leq \text{vol}(g S_{H_0, 2r} \cap S_{H_0, r})$$

for all  $g \in S_{H,r}$ . To see this, we first note that by the previous paragraph we have that  $\text{vol}(e^H S_{H_0, r} \cap S_{H_0, r}) \leq \text{vol}(e^{H+Z} S_{H_0, 2r} \cap S_{H_0, r})$  for any  $Z \in B_{\mathfrak{a}}(0, r)$ . Any element  $g \in S_{H,r}$  can be expressed as  $ke^{H+Z}k'$  for some  $k, k' \in K$  and  $Z \in B_{\mathfrak{a}}(0, r)$ . However, the sets  $S_{H_0, r}$  are bi- $K$ -invariant, so  $\text{vol}(g S_{H_0, 2r} \cap S_{H_0, r}) = \text{vol}(e^{H+Z} S_{H_0, 2r} \cap S_{H_0, r})$ .

We can re-express (48) as

$$\text{vol}(e^H S_{H_0, r} \cap S_{H_0, r}) \mathbb{1}_{S_{H_0, r}} \leq \mathbb{1}_{S_{H_0, r}} * \mathbb{1}_{S_{H_0, 2r}}^\vee.$$

We now apply the endpoint Kunze–Stein property to obtain:

$$\text{vol}(e^H S_{H_0, r} \cap S_{H_0, r}) \|\mathbb{1}_{S_{H_0, r}}\|_{2, \infty} \leq \|\mathbb{1}_{S_{H_0, r}}\|_{2, 1} \|\mathbb{1}_{S_{H_0, 2r}}\|_{2, 1}.$$

Finally we have the elementary estimate

$$\|\mathbb{1}_{S_{H_0, r}}\|_{2, q} \asymp e^{\rho(H_0)}$$

for all  $1 \leq q \leq \infty$ . This implies  $\text{vol}(e^H S_{H_0, r} \cap S_{H_0, r}) \ll e^{\rho(2H_0 - H)}$ .

**6.3. Reduction to a spectral estimate.** We now return to the usual notation for  $S_t$ , i.e., as defined in (27). Let  $\psi \in C_c^\infty(\mathfrak{a})$  be a non-negative  $W_M$ -invariant function such that  $\psi \geq 1$  on  $B_{\mathfrak{a}}(0, \epsilon_0)$ . Let  $\psi_t$  be the translate of  $\psi$  by  $tH_0$ , given by  $\psi_t(Z) = \psi(Z - tH_0)$ , which satisfies  $\psi_t \geq 1$  on  $B_{\mathfrak{a}}(tH_0, \epsilon_0)$ . We then define  $k_t \in C_c^\infty(G)$  to be the bi- $K$ -invariant function that satisfies  $k_t(e^H) = \sum_{w \in W/W_M} \psi_t(H)$ , which likewise has the property that  $k_t \geq 1$  on  $S_t$ .

We have  $\text{vol}(S_t \cap e^H S_t) \leq (k_t * k_t^\vee)(e^H)$ , so that it suffices to estimate the convolution  $k_t * k_t^\vee$ . We shall do this using the Harish-Chandra transform, by writing

$$(49) \quad (k_t * k_t^\vee)(e^H) = \int_{\mathfrak{a}_+^*} |\widehat{k}_t(\lambda)|^2 \varphi_\lambda(e^H) |c(\lambda)|^{-2} d\lambda.$$

We shall bound this expression using Theorem 2.3, together with the following result which is obtained from that theorem through a uniform integration by parts argument. We recall the function  $\Theta(H, \lambda)$  appearing in the statement of Theorem 2.3, defined in (4).

**Proposition 6.1.** *Let  $G$  be a non-compact semisimple Lie group with finite center. Let  $H_0 \in \bar{\mathfrak{a}}_+$ . For any  $N > 0$  we have*

$$\widehat{k}_t(\lambda) \ll_N (1 + \|\lambda\|)^{-N} e^{\rho(tH_0)} \Theta(tH_0, \lambda),$$

uniformly in  $t$ .

Applying Theorem 2.3 and Proposition 6.1 in (49), we obtain

$$(k_t * k_t^\vee)(e^H) \ll_N e^{\rho(2tH_0 - H)} \int_{\mathfrak{a}_+^*} (1 + \|\lambda\|)^{-N} \Theta(tH_0, \lambda)^2 \Theta(H, \lambda) |c(\lambda)|^{-2} d\lambda.$$

Note that  $(k_t * k_t^\vee)(e^H) = 0$  if  $t \ll \|H\|$ . We must therefore prove the following estimate of the elementary spectral integral.

**Proposition 6.2.** *Let  $G$  be a non-compact simple real Lie group. If  $H_0 \in \bar{\mathfrak{a}}_+$  is such that  $\Phi_{M, \text{red}}$  is semi-dense in  $\Phi_{\text{red}}$ , there exists an integer  $k \in \mathbb{Z}_{\geq 0}$  such that*

$$(50) \quad \int_{\mathfrak{a}_+^*} (1 + \|\lambda\|)^{-N} \Theta(tH_0, \lambda)^2 \Theta(H, \lambda) |c(\lambda)|^{-2} d\lambda \ll (\log t)^k$$

for  $\|H\| \ll t$ .

The combination of Theorem 2.3 and Propositions 6.1-6.2 then implies Theorem 2.2. The remainder of the paper is dedicated to proving these three key ingredients.

**6.4. Proof of Proposition 6.1.** We let  $\Delta$  denote the Laplacian on the symmetric space  $X = G/K$ . Recall that for  $\lambda \in \mathfrak{a}^*$  we have  $\Delta\varphi_\lambda = (\|\rho\|^2 + \|\lambda\|^2)\varphi_\lambda$ , see [Hel00, p.427, (7)]. Let  $N$  be a positive integer. Integrating by parts (using Green's second identity)  $N$  times we obtain

$$\widehat{k}_t(\lambda) = (\|\rho\|^2 + \|\lambda\|^2)^{-N} \int_X k_t(x) \Delta^N \varphi_{-\lambda}(x) dx = (\|\rho\|^2 + \|\lambda\|^2)^{-N} \int_X \Delta^N k_t(x) \varphi_{-\lambda}(x) dg.$$

Here we have identified the right  $K$ -invariant functions  $k_t(g)$  and  $\phi_{-\lambda}(g)$  with the corresponding functions on  $X$ , which is the space on which  $\Delta$  is defined. Since the integrand is also left  $K$ -invariant we can use the integral decomposition (11) to write

$$(51) \quad \widehat{k}_t(\lambda) = \textcolor{violet}{b}_G (\|\rho\|^2 + \|\lambda\|^2)^{-N} \int_{\mathfrak{a}_+} \Delta^N k_t(e^H) \varphi_{-\lambda}(e^H) J(H) dH.$$

Proposition 6.1 will follow from the following bound on  $\Delta^N k_t$ .

**Lemma 6.3.** *Let  $k_t$  be as defined in §6.3. Let  $B \subset \mathfrak{a}$  be a  $W_M$ -invariant compact subset such that  $\text{supp}(\psi) \subset B$ . Then, for all  $H \in \mathfrak{a}$  and all  $N \geq 0$ ,*

$$\Delta^N k_t(e^H) \ll_N \sum_{w \in W/W_M} \mathbf{1}_B(H - wtH_0),$$

uniformly in  $t > 0$ .

To see how Proposition 6.1 follows from this, if we insert Lemma 6.3 and Theorem 2.3 into (51), we find

$$\widehat{k}_t(\lambda) \ll_N (1 + \|\lambda\|)^{-N} \int_{(tH_0 + B) \cap \mathfrak{a}_+} e^{-\rho(H)} J(H) \Theta(H, \lambda) dH.$$

For  $t$  large, the integrand is roughly constant of size  $e^{\rho(tH_0)} \Theta(tH_0, \lambda)$ , yielding the statement of Proposition 6.1

We now turn to proving Lemma 6.3. We will prove this by working in radial coordinates. Let  $x_0 \in X = G/K$  be the point corresponding to the trivial coset  $K$ . Let  $\Delta_{\text{rad}}$  denote the radial part of  $\Delta$  relative to the  $K$ -action on  $X$  with  $A_+ \cdot x_0$  as a transversal manifold, where  $A_+ = \exp(\mathfrak{a}_+)$ . Then by the bi- $K$ -invariance of  $k_t$  we have  $\Delta k_t(e^H) = \Delta_{\text{rad}} k_t(e^H)$ .

We recall the expression for  $\Delta_{\text{rad}}$  from Ch. II, Proposition 3.9 of [Hel00]. Let  $\Delta_{\mathfrak{a}}$  denote the Laplacian on the flat manifold  $A \cdot x_0 \subset X$ , which is isometric to  $\mathfrak{a}$ . Then

$$\Delta_{\text{rad}} = \Delta_{\mathfrak{a}} + \sum_{\alpha \in \Phi^+} m_\alpha(\coth \alpha) X_\alpha,$$

where  $m_\alpha = \dim \mathfrak{g}_\alpha$ , and the element  $X_\alpha \in \mathfrak{a}$ , defined by  $\langle X_\alpha, H \rangle = \alpha(H)$ , is viewed as a differential operator on  $A \cdot x_0$ . Instead of working on  $A \cdot x_0$ , we can equivalently work on  $\mathfrak{a}$  instead, and shall thus instead view  $\Delta_{\mathfrak{a}}$  and  $X_\alpha$  as differential operators on  $\mathfrak{a}$ .

We define the translated operator  $\Delta_{\text{rad}}^t = T_{-tH_0} \Delta_{\text{rad}} T_{tH_0}$ , where  $T_H$  denotes the operator of translation by  $H$ . Explicitly,  $\Delta_{\text{rad}}^t$  is given by

$$\Delta_{\text{rad}}^t = \Delta_{\mathfrak{a}} + \sum_{\alpha \in \Phi^+} m_\alpha \coth(\alpha(\cdot - tH_0)) X_\alpha.$$

Thus

$$k_t(e^H) = \sum_{w \in W/W_M} \psi(wH - tH_0) = \sum_{w \in W/W_M} T_{tH_0} \psi(wH).$$

Therefore,

$$\Delta^N k_t(e^H) = \sum_{w \in W/W_M} (T_{tH_0}(\Delta_{\text{rad}}^t)^N \psi)(wH) = \sum_{w \in W/W_M} ((\Delta_{\text{rad}}^t)^N \psi)(wH - tH_0).$$

It therefore suffices to prove that  $\Delta_{\text{rad}}^{t,N} \psi \ll_N \mathbb{1}_B$  uniformly in  $t$ . This follows from the following lemma.

**Lemma 6.4.** *Suppose  $\mathcal{B}$  is a bounded subset of  $C_c^\infty(\mathfrak{a})^{W_M}$  whose elements are supported in a compact set  $\mathcal{K}$ . If  $t_0 > 0$  is such that the functions  $H \mapsto \alpha(H - tH_0)$  are all non-vanishing on  $\mathcal{K}$  for  $t \geq t_0$  and  $\alpha \in \Phi^+ \setminus \Phi_M^+$ , then the set  $\bigcup_{t \geq t_0} \Delta_{\text{rad}}^t \mathcal{B}$  is also bounded in  $C_c^\infty(\mathfrak{a})^{W_M}$ .*

*Proof.* The operator  $\Delta_{\text{rad}}^t$  has singularities along the root hyperplanes corresponding to roots in  $\Phi_M^+$ . To deal with these, it will be convenient to also introduce the symmetric space  $X_M = M/K_M$ , with Laplacian  $\Delta_M$ . The radial part of  $\Delta_M$  is given by

$$\Delta_{M,\text{rad}} = \Delta_{\mathfrak{a}} + \sum_{\alpha \in \Phi_M^+} m_\alpha \coth(\alpha) X_\alpha,$$

so that

$$\Delta_{\text{rad}}^t = \Delta_{M,\text{rad}} + \sum_{\alpha \in \Phi^+ \setminus \Phi_M^+} m_\alpha \coth(\alpha(\cdot - tH_0)) X_\alpha.$$

The advantage of doing this is that we have written  $\Delta_{\text{rad}}^t$  as the sum of the radial part of the smooth operator  $\Delta_M$ , and the operators  $\coth(\alpha(\cdot - tH_0)) X_\alpha$  for  $\alpha \in \Phi^+ \setminus \Phi_M^+$ , which are non-singular on functions supported on any fixed compact set satisfying the hypotheses of the lemma.

We shall prove that  $\Delta_{M,\text{rad}}$  is bounded on  $C_c^\infty(\mathfrak{a})^{W_M}$  by reducing to the boundedness of  $\Delta_M$ . To do this, we define the restriction operator  $R : C_c^\infty(X_M)^{K_M} \rightarrow C_c^\infty(\mathfrak{a})^{W_M}$ . Helgason [Hel00] proves in Ch. II, Section 5, Theorem 5.8 that  $R$  is an isomorphism of topological vector spaces, and therefore has a continuous inverse which we denote by  $E : C_c^\infty(\mathfrak{a})^{W_M} \rightarrow C_c^\infty(X_M)^{K_M}$ .

We have  $\Delta_{M,\text{rad}} = R \circ \Delta_M \circ E$ , and as the operators  $E$ ,  $R$ , and  $\Delta_M$  are all bounded on  $C_c^\infty(\mathfrak{a})^{W_M}$  or  $C_c^\infty(X_M)^{K_M}$ , we see that  $\Delta_{M,\text{rad}}$  is also bounded on  $C_c^\infty(\mathfrak{a})^{W_M}$ , and thus  $\bigcup_{t \geq t_0} \Delta_{M,\text{rad}} \mathcal{B}$  is a bounded set.

For any choice of  $\varepsilon > 0$ , we have that for every  $n$  there exists a  $C_n$  such that  $|\frac{d^n}{dx^n} \coth(x)| < C_n$  for  $x \in (\varepsilon, \infty)$ . Therefore, by the product rule, we conclude that

$$\bigcup_{t \geq t_0} \left( \sum_{\alpha \in \Phi^+ \setminus \Phi_M^+} m_\alpha \coth(\alpha(\cdot - tH_0)) X_\alpha \right) \mathcal{B}$$

is also a bounded set. □

## 7. SEMI-DENSE ROOT SUBSYSTEMS

We begin by introducing a property of root subsystems, which will be a key element in the proof of the bound on intersection volumes in Theorem 2.2.

Throughout this section we suppose  $\Phi$  is a reduced root system of rank  $n$  in a vector space  $V$  with  $\Delta$  as a base of simple roots. We say that a root subsystem of  $\Phi$  is *standard* if it is the root subsystem generated by a subset of the simple roots, and we say that it is *semistandard* if it is the image of a standard root subsystem under an element of the Weyl group. Semistandard root subsystems can be characterized as those arising as  $\Phi \cap V'$  where  $V'$  is some subspace of  $V$  (Chapter VI, §1, Proposition 24 of [Bou02]).

**Definition 7.1.** *Let  $\Phi$  be an reduced root system. We say that a semistandard root subsystem  $\Phi_0 \subset \Phi$  is semi-dense if, for every semistandard root subsystem  $\Psi \subset \Phi$  we have*

$$(52) \quad |\Psi \cap \Phi_0| + \text{rank}(\Psi) \geq \frac{1}{2}|\Psi|.$$

*Remark 7.2.* Suppose  $\Phi = \Phi^{(1)} \sqcup \dots \sqcup \Phi^{(m)}$  with each  $\Phi^{(j)}$  irreducible. Suppose  $\Phi^{(1)}$  admits a semi-dense root subsystem  $\Phi_0^{(1)}$ . Then the subroot system  $\Phi_0^{(1)} \sqcup \Phi^{(2)} \sqcup \dots \sqcup \Phi^{(m)}$  is semi-dense in  $\Phi$ .

In this section, we will show that the root system corresponding to an extremal element is semi-dense in the above sense (Proposition 7.5) and that no root subsystem in types  $E_6, E_8, F_4, G_2$  is semi-dense (Lemma 7.7). The latter negative result accounts for the exclusion of such root systems in Theorem 1.1.

*Remark 7.3.* We emphasize, however, that classical root systems can admit semi-dense root subsystems not arising from extremal elements. Most notably, this is the case for maximal semistandard root subsystems  $\Phi_0 \subset \Phi$ , when  $\Phi$  is type  $A_n$ . The class of root subsystems arising from extremal elements is a convenient choice though because the corresponding semi-dense root subsystem minimizes how many  $\Psi$ 's achieve equality in (52), which in turn will minimize the power  $k$  of  $\log t$  appearing in Theorem 5.1 (see Lemma 8.8). Furthermore, they have the biggest Weyl group which ultimately minimizes the number of hyperplanes to be avoided in the spectral bound (Theorem 4.1).

**7.1. Classical root systems.** We first focus on root systems of classical type, i.e.,  $A_n, B_n, C_n$  and  $D_n$ .

**Definition 7.4.** *Let  $\Phi$  be a rank  $n$  root system of classical type. We say that a root subsystem  $\Phi_0$  of  $\Phi$  is extremal if it is of rank  $n-1$ , irreducible, semistandard, and in the same family as  $\Phi$  (with the conventions that  $B_1 := A_1$ ,  $C_2 := B_2$ , and  $D_3 := A_3$ ). All such root subsystems arise by first taking a standard extremal root subsystem and then applying to it some element of  $W$ .*

It is clear that extremal root subsystems are simply root subsystems obtained by applying an element of  $W$  to a root system corresponding to an extremal element in the sense of Section 3.6.

**Proposition 7.5.** *Suppose  $\Phi$  is an irreducible root system of classical type. Then any extremal root subsystem is semi-dense.*

*Proof.* Let  $\Phi_0 \subset \Phi$  be extremal and let  $\Psi \subset \Phi$  be semistandard. In fact, it will be enough to verify Definition 7.1 with  $\Psi$  standard. Indeed, writing  $\Psi = w\Psi_{\text{st}}$ , where  $\Psi_{\text{st}} \subset \Phi$  is a standard root subsystem, we have  $|\Psi \cap \Phi_0| = |\Psi_{\text{st}} \cap w^{-1}\Phi_0|$ . Since  $\text{rank}(\Psi) = \text{rank}(\Psi_{\text{st}})$  and  $|w^{-1}\Phi_0| = |\Phi_0|$ , inequality (52) is true for  $\Psi$  if and only if it is true for  $\Psi_{\text{st}}$ .

We now let  $\Psi$  be an arbitrary standard root subsystem of  $\Phi$ . To prove the proposition, it is enough to establish the inequality (52) for each of the irreducible factors of  $\Psi$ . We may therefore assume that  $\Psi$  is irreducible. Moreover, we may assume that  $\Psi \not\subset \Phi_0$  since otherwise the inequality (52) is obviously true:  $|\Psi| + \text{rank}(\Psi) \geq |\Psi|/2$ . We shall assume throughout the remainder of this proof that  $\Psi$  is *standard, irreducible and not strictly contained in  $\Phi_0$* .

Our strategy is to reduce the verification of (52) to the base case  $\Psi = \Phi$ , which can be checked numerically. In this case we must verify

$$(53) \quad |\Phi_0| + \text{rank}(\Phi) \geq \frac{1}{2}|\Phi|,$$

which can be checked by inspection of the following table:

type of $\Phi$	type of $\Phi_0$	$ \Phi $	$ \Phi_0 $	(53) true?
$A_n$	$A_{n-1}$	$n(n+1)$	$(n-1)n$	yes
$B_n$	$B_{n-1}$	$2n^2$	$2(n-1)^2$	yes
$C_n$	$C_{n-1}$	$2n^2$	$2(n-1)^2$	yes
$D_n$	$D_{n-1}$	$2n(n-1)$	$2(n-1)(n-2)$	yes

It remains then to show that when  $\Psi \neq \Phi$  then  $\Psi \cap \Phi_0$  is extremal in  $\Psi$ , in which case we can apply the base case to conclude. (We will in fact encounter the degenerate setting when  $\Psi$  is a singleton set, in which case  $\Psi$  is of type  $A_1$ ,  $\Psi \cap \Phi_0 = \emptyset$ , and the inequality (52) is trivially verified, since  $0 + 1 \geq \frac{1}{2} \cdot 2$ .) It will be enough to prove this claim for  $\Psi$  maximal. Note that maximal standard subsystems correspond to removing one simple root from  $\Delta$ .

We divide the proof of the above claim according to the root type.

*Type  $A_n$ .* The root system of type  $A_n$  has a model as vectors in  $\mathbb{R}^{n+1}$  of the form  $e_i - e_j$  with  $i \neq j$ . Let  $V$  be the subspace spanned by the roots (i.e., the hyperplane orthogonal to  $e_1 + \dots + e_{n+1}$ ). A basis of simple roots is  $e_1 - e_2, \dots, e_n - e_{n+1}$ . Let  $\Phi'_0$  be the extremal root subsystem generated by  $e_2 - e_3, \dots, e_n - e_{n+1}$ . This is exactly the intersection of  $\Phi$  with the hyperplane orthogonal to  $e_1$ . All extremal root subsystems are of the form  $\Phi_0 = w\Phi'_0$ . There are precisely  $n+1$  of these each of the form  $\Phi \cap \{\langle e_j, x \rangle = 0\}$  for some  $1 \leq j \leq n+1$ .

Suppose  $\Phi_0 = \Phi \cap \{\langle e_j, x \rangle = 0\}$ . We wish to show that  $\Phi_0 \cap \Psi$  is extremal in  $\Psi$ . For  $1 \leq l < n+1$  let  $V_l^{n+1}$  and  $W_l^{n+1}$  denote the linear subspaces of  $\mathbb{R}^{n+1}$  generated by  $e_1, \dots, e_l$  and  $e_{l+1}, \dots, e_{n+1}$ , respectively. Then  $\Psi = (\Psi \cap V_l^{n+1}) \times (\Psi \cap W_l^{n+1}) =: \Psi_1 \times \Psi_2$  for a suitable  $1 \leq l < n+1$ . Then  $\Psi_1$  and  $\Psi_2$  are of type  $A_{l-1}$  and  $A_{n-l}$ , respectively. Moreover,  $\Phi_0 \cap \Psi$  equals either  $\Psi_{1,0} \times \Psi_2$  or  $\Psi_1 \times \Psi_{2,0}$ , depending on whether  $j \leq l$  or  $j \geq l+1$ , where  $\Psi_{k,0} = \Psi_k \cap \{\langle e_j, x \rangle = 0\}$ . By above discussion,  $\Psi_{k,0}$  is extremal in  $\Psi_k$ , and thus  $\Phi_0 \cap \Psi$  is extremal in  $\Psi$ .

There are at most two simple roots in  $\Phi$  which are not orthogonal to  $e_j$ , namely  $e_{j-1} - e_j$  and  $e_j - e_{j+1}$  (if  $j = 1$  or  $n+1$ , then there is only one simple root not orthogonal to  $e_j$ ). Since  $\Psi$  is not contained in  $\Phi_0$ , it contains a simple root not orthogonal to  $e_j$ . Without loss of generality, we may suppose that  $\Phi_0 \cap \Psi$  is generated by the simple roots  $e_k - e_{k+1}, \dots, e_{k+\ell-1} - e_{k+\ell}$  with  $k \leq j \leq k+\ell$  (and in particular is of type  $A_\ell$ ). If  $j = k$  or  $k+\ell$ , then it is clear

that  $\Phi_0 \cap \Psi$  is extremal in  $\Psi$ . Otherwise, we can instead take as our basis for  $\Psi$  the set  $e_j - e_k, e_k - e_{k+1}, \dots, e_{j-1} - e_{j+1}, e_{j+1} - e_{j+2}, \dots, e_{k+\ell-1} - e_{k+\ell}$ . Then  $\Phi_0 \cap \Psi$  is the root subsystem obtained by removing  $e_j - e_k$  from the above basis. Thus the intersection is extremal (and in particular it is of type  $A_{\ell-1}$ ).

*Types  $B_n$  and  $C_n$ .* The  $B_n$  and  $C_n$  cases are virtually identical, and thus we just discuss  $B_n$ . The analysis is in turn very similar to the  $A_n$  case. We have as a model of our root system vectors of the form  $\pm e_i \pm e_j$  with  $i \neq j$  and  $\pm e_j$  inside of  $\mathbb{R}^n$ . We may take as basis  $e_1 - e_2, e_2 - e_3, \dots, e_{n-1} - e_n, e_n$ . Let  $\Phi'_0$  be the extremal root subsystem generated by the  $e_2 - e_3, \dots, e_{n-1} - e_n, e_n$ . This is exactly the intersection of  $\Phi$  with the hyperplane orthogonal to  $e_1$ . All other extremal root subsystems are of the form  $\Phi_0 = w\Phi'_0$ . There are exactly  $n$  of these all of the form  $\Phi \cap \{\langle e_j, x \rangle = 0\}$  for some  $1 \leq j \leq n$  (because the orbit of  $e_1$  under  $W$  is exactly vectors of the form  $\pm e_j$ ).

Suppose  $\Phi_0 = \Phi \cap \{\langle e_j, x \rangle = 0\}$ . If  $\Psi$  is obtained by removing one of the first  $n-1$  simple roots, we again have  $\Psi = (\Psi \cap V_l^n) \times (\Psi \cap W_l^n) =: \Psi_1 \times \Psi_2$  for some  $1 \leq l < n$ , but with  $\Psi_2$  now of type  $B_{n-l}$  or  $C_{n-l}$ . The claim then follows as in the  $A_n$  case. If  $\Psi$  is obtained by removing  $e_n$ , then  $\Psi$  is the usual  $A_{n-1}$  root system in  $\mathbb{R}^n$  and we are also back in the previous case.

If  $j = 1$  then  $e_1 - e_2$  is the only simple root not orthogonal to  $e_j$ . If  $1 \leq j \leq n-1$ , then  $e_{j-1} - e_j$  and  $e_j - e_{j+1}$  are the only simple roots not orthogonal to  $e_j$ . If  $j = n$ , then  $e_{n-1} - e_n$  and  $e_n$  are the only simple roots not orthogonal to  $e_j$ . Since  $\Psi$  is not contained in  $\Phi_0$ , it contains a simple root not orthogonal to  $e_j$ . Thus, either  $\Psi$  is of type  $A_\ell$  in which case the analysis is exactly the same as before, or it is of type  $B_\ell$ , namely we have  $e_k - e_{k+1}, \dots, e_{n-1} - e_n, e_n$ . If  $j \neq n$ , then we can perform the same trick as before, i.e., choose the basis  $e_j - e_k, e_k - e_{k+1}, \dots, e_{j-1} - e_{j+1}, \dots, e_{n-1} - e_n, e_n$  to conclude that the intersection is extremal (in particular of type  $B_{\ell-1}$ ). Otherwise suppose  $j = n$ . Then we take as our basis:  $e_n - e_k, e_k - e_{k+1}, \dots, e_{n-2} - e_{n-1}, e_{n-1}$  in which case we again see that the intersection is extremal.

*Type  $D_n$ .* The type  $D_n$  root system has a model as vectors in  $\mathbb{R}^n$  of the form  $\pm e_i \pm e_j$  with  $i \neq j$ . We may take  $\Delta = \{e_1 - e_2, e_2 - e_3, \dots, e_{n-1} - e_n, e_{n-1} + e_n\}$ . Let  $\Phi'_0$  be the standard extremal root subsystem obtained by removing  $e_1 - e_2$  from  $\Delta$ . Then  $\Phi'_0 = \Phi \cap \{\langle e_1, x \rangle = 0\}$ . All other extremal root subsystems are of the form  $\Phi_0 = w\Phi'_0$ ; each such one can be expressed as  $\Phi \cap \{\langle e_j, x \rangle = 0\}$  for some  $j$ .

Suppose  $\Phi_0 = \Phi \cap \{\langle e_j, x \rangle = 0\}$ . If  $\Psi$  is obtained by deleting one of the first  $n-2$  simple roots, then we proceed as before by looking at the intersection of  $\Psi$  with  $V_l^n$  and  $W_l^n$  (now  $\Psi_2$  is of type  $D_{n-l}$ ). If  $\Psi$  is obtained by deleting the last simple root,  $\Psi$  is just the usual  $A_{n-1}$  in  $\mathbb{R}^n$  and we are back in the first case. Finally, if  $\Psi$  is obtained by removing the second to last root in  $\Delta$ , then  $\Psi$  is also of type  $A_{n-1}$  but with basis  $e_1 - e_2, \dots, e_{n-2} - e_{n-1}, e_{n-1} + e_n$ . But since  $\{\langle e_n, x \rangle = 0\} = \{\langle -e_n, x \rangle = 0\}$ , we are again back in the first case.

If  $j = 1$ , then  $e_1 - e_2$  is the only simple root not orthogonal to  $e_1$ . If  $2 \leq j \leq n-2$ , then there are exactly two simple roots not orthogonal to  $e_j$ . If  $j = n-1$ , there are three simple roots not orthogonal to  $e_j$ , namely  $e_{n-2} - e_{n-1}, e_{n-1} - e_n, e_{n-1} + e_n$ . If  $j = n$ , there are two simple roots not orthogonal to  $e_j$ :  $e_{n-1} - e_n, e_{n-1} + e_n$ . By the same style of analysis as before, it is clear that if  $1 \leq j \leq n-2$ , then we are done.

Now suppose  $j = n$ . Suppose first that  $\Psi$  does not contain the simple root  $e_{n-2} - e_{n-1}$ . Since  $\Psi$  is irreducible and is not contained in  $\Phi_0$ , it follows that  $\Psi$  is the singleton set

consisting of either  $e_{n-1} - e_n$  or  $e_{n-1} + e_n$ . This is the degenerate setting referred to earlier, in which (52) reduces to  $0 + 1 \geq \frac{1}{2} \cdot 2$ . Otherwise,  $\Psi$  contains  $e_{n-2} - e_{n-1}$ , in which case  $\Psi$  contains one or both of  $e_{n-1} - e_n$  and  $e_{n-1} + e_n$ . If  $\Psi$  contains only one of these extremal roots, then  $\Psi$  is of type  $A_\ell$ , and the same style of analysis as before shows that  $\Psi \cap \Phi_0$  is extremal in  $\Psi$ . Otherwise  $\Psi$  contains both  $e_{n-1} - e_n$  and  $e_{n-1} + e_n$ . In this case  $\Psi$  is of type  $D_\ell$  with corresponding simple roots  $e_k - e_{k+1}, \dots, e_{n-2} - e_{n-1}, e_{n-1} - e_n, e_{n-1} + e_n$  (with  $\ell = n - k + 1$ ). We can instead choose the basis  $e_n - e_k, e_k - e_{k+1}, \dots, e_{n-3} - e_{n-2}, e_{n-2} - e_{n-1}, e_{n-2} + e_{n-1}$ . It is then clear that  $\Psi \cap \Phi_0$  is extremal in  $\Psi$  (i.e., it is of type  $D_{\ell-1}$ , with the convention that  $D_2 = A_1 \times A_1$ ).

Finally suppose  $j = n - 1$ . Suppose first that  $\Psi$  does not contain the simple root  $e_{n-2} - e_{n-1}$ . Then, like the  $j = n$  case,  $\Psi$  is the singleton set consisting of either  $e_{n-1} - e_n$  or  $e_{n-1} + e_n$ , in which case we conclude as before. Otherwise,  $\Psi$  contains  $e_{n-2} - e_{n-1}$ , in which case  $\Psi$  either contains exactly one of  $e_{n-1} - e_n$  and  $e_{n-1} + e_n$  or contains both. In the former case, we are again back in the type  $A$  case. Otherwise,  $\Psi$  contains both  $e_{n-1} - e_n$  and  $e_{n-1} + e_n$  and is of type  $D_\ell$  with corresponding basis  $e_k - e_{k+1}, \dots, e_{n-2} - e_{n-1}, e_{n-1} - e_n, e_{n-1} + e_n$ . We instead take the basis  $e_{n-1} - e_{k+1}, e_{k+1} - e_{k+2}, \dots, e_{n-2} - e_k, e_k - e_n, e_k + e_n$ . It is then clear that  $\Psi \cap \Phi_0$  is extremal in  $\Psi$  (again with the convention that  $D_2 = A_1 \times A_1$ ).  $\square$

*Remark 7.6.* We shall show an example of a classical root system  $\Phi$  which shows that, if we do not take  $\Psi$  to be extremal, then (52) might not hold. Let  $\Phi$  be of type  $B_3$ , and let  $\Psi$  be the standard  $A_1 \times A_1$  root subsystem. Then  $\Phi$  has 18 roots, and  $\Psi$  has 4 roots. We thus see that  $|\Phi \cap \Psi| + \text{rank}(\Phi) = 4 + 3$ , but this is smaller than  $\frac{1}{2}|\Phi| = 9$ .

## 7.2. Exceptional root systems.

**Lemma 7.7.** *Suppose  $\Phi$  is of type  $E_6, E_8, F_4$  or  $G_2$ . Then there exists no semi-dense root subsystem.*

*Proof.* In the case of  $G_2$  this is immediate: the only root subsystem is of type  $A_1$ , and if we take  $\Psi = \Phi$  we see that the inequality fails ( $G_2$  has 12 roots).

The case of  $F_4$  is also straightforward: one may check that if we take  $\Psi = \Phi$  (which has 48 roots) and  $\Phi_0$  to be any standard proper root subsystem, then the desired inequality fails.

The type  $E$  cases are more subtle and the desired inequality “just barely” fails to be true (and in fact it holds for  $E_7$ !). If we take for  $\Phi_0$  the  $E_7$  standard root subsystem in  $E_8$ , or the  $D_5$  standard root subsystem in  $E_6$ , then the desired inequality in fact holds for any  $\Psi$  standard. Thus constructing counterexamples to (52) is more complicated.

We start with the case of  $\Phi$  being the  $E_8$  root system. There are three extremal nodes of the Dynkin diagram of  $\Phi$ ; removing these result in root subsystems of types  $E_7, A_7$ , and  $D_7$ . Recall that  $E_8$  has 240 roots. If we take  $\Phi_0$  to be of type  $D_7$  (which has 84 roots) or type  $A_7$  (which has 56 roots), then the desired inequality fails if we simply take  $\Psi = \Phi$ . Similarly, if we take  $\Phi_0$  to be any other standard root subsystem other than type  $E_7$  (which has 126 roots), then the desired inequality fails. So our only chance is to take  $\Phi_0$  to be of type  $E_7$ .

We can take as our model for  $\Phi$  the elements in  $\mathbb{R}^8$  of the form  $\pm e_i \pm e_j$  with  $i \neq j$  as well as  $\pm \frac{1}{2}e_1 \pm \frac{1}{2}e_2 \pm \dots \pm \frac{1}{2}e_8$  with an even number of minus signs (sum of coefficients is even).

We can take as our basis of simple roots:

$$\begin{aligned} & \{e_1 - e_2, e_2 - e_3, e_3 - e_4, e_4 - e_5, e_5 - e_6, e_6 - e_7, e_6 + e_7, \\ & - \frac{1}{2}(e_1 + e_2 + e_3 + e_4 + e_5 + e_6 + e_7 + e_8)\}. \end{aligned}$$

Let  $\Phi_0$  be the  $E_7$  standard root subsystem which is obtained from removing  $e_1 - e_2$  from the base. This is exactly the intersection of  $\Phi$  with the hyperplane orthogonal to  $e_1 - e_8$ .

Let  $w \in W$  be the element switching  $e_3$  and  $e_8$ . Then  $w\Phi_0$  consists of roots orthogonal to  $e_1 - e_3$ . Consider the standard root subsystem  $\Psi$  generated by  $e_1 - e_2$  and  $e_2 - e_3$ . This generates an  $A_2$  type root system. However, no root in this root subsystem is orthogonal to  $e_1 - e_3$ . Therefore  $w\Phi_0 \cap \Psi = \emptyset$ . Thus  $|w\Phi_0 \cap \Psi| + \text{rank}(\Psi) = 2$  but  $\frac{1}{2}|\Psi| = 3$ . Thus the desired inequality does not hold for  $E_8$ .

Now suppose  $\Phi$  is of type  $E_6$ . Recall that  $E_6$  has 72 roots. Removing an extremal node from the Dynkin diagram gives root subsystems of type  $D_5$  (with 40 roots) and  $A_5$  (with 30 roots). Both of these root subsystems in fact satisfy (52) in case  $\Psi = \Phi$ . However, any other root subsystem other will fail to satisfy (52) by again simply taking  $\Psi = \Phi$ . Thus we must only analyze further the cases of  $A_5$  and  $D_5$ .

Recall that the  $A_2$  root system has a model as vectors in  $\mathbb{R}^3$  of the form  $(1, -1, 0)$ ,  $(1, 0, -1)$ ,  $(0, 1, -1)$  or their negatives. We may take as our model of  $E_6$  those vectors in  $\mathbb{R}^9 = \mathbb{R}^3 \times \mathbb{R}^3 \times \mathbb{R}^3$  where we separately take a copy of  $A_2$  inside of each of three copies of  $\mathbb{R}^3$ , together with vectors of the form  $(A; B; C)$  where  $A, B, C \in \{(\frac{2}{3}, -\frac{1}{3}, -\frac{1}{3}), (-\frac{1}{3}, \frac{2}{3}, -\frac{1}{3}), (-\frac{1}{3}, -\frac{1}{3}, \frac{2}{3})\}$ , and vectors of the form  $(A'; B'; C')$  where  $A', B', C' \in \{(-\frac{2}{3}, \frac{1}{3}, \frac{1}{3}), (\frac{1}{3}, -\frac{2}{3}, \frac{1}{3}), (\frac{1}{3}, \frac{1}{3}, -\frac{2}{3})\}$ . We can take as our simple roots

$$\begin{aligned} & \{e_8 - e_9, e_7 - e_8, e_5 - e_6, e_4 - e_5, e_2 - e_3, \\ & \frac{1}{3}e_1 - \frac{2}{3}e_2 + \frac{1}{3}e_3 - \frac{2}{3}e_4 + \frac{1}{3}e_5 + \frac{1}{3}e_6 - \frac{2}{3}e_7 + \frac{1}{3}e_8 + \frac{1}{3}e_9\}. \end{aligned}$$

These roots are all orthogonal to  $e_1 + e_2 + e_3$ ,  $e_4 + e_5 + e_6$ , and  $e_7 + e_8 + e_9$ .

First we let  $\Phi_0$  be the standard root system of type  $A_5$ . This is obtained taking the root subsystem generated by the simple roots with  $e_2 - e_3$  removed; equivalently these are the roots orthogonal to  $e_1 - e_3$ . Let  $w \in W$  be the element switching  $e_1$  and  $e_2$ . Then  $e_1 - e_3$  is moved to  $e_2 - e_3$ , and  $w\Phi_0$  is those roots orthogonal to  $e_2 - e_3$ . Let  $\Psi$  be the root subsystem generated by the last two simple roots in the list above. Then  $\Psi$  is of type  $A_2$ . However no vector in  $\Psi$  is orthogonal to  $e_2 - e_3$ . Therefore  $|w\Phi_0 \cap \Psi| = 0$ , and we see that (52) fails.

Lastly we consider  $\Phi_0$  of type  $D_5$ . We shall use a different model for  $E_6$  than the one above. Specifically, we shall take  $\Phi$  to be the root subsystem of  $E_8$  obtained from removing the simple roots  $e_1 - e_2$  and  $e_2 - e_3$  from the base of  $E_8$ . Then  $\Phi$  is the root subsystem orthogonal to  $e_1 - e_2$  and  $e_1 - e_8$ . Let  $\Phi_0$  be the root subsystem of  $\Phi$  obtained by further removing the element  $e_3 - e_4$  from the base. These are the roots which lie in the intersection of the space containing  $\Phi$  with the hyperplane orthogonal to  $e_1 + e_2 - 3e_3 + e_8$ . Let  $w \in W$  be the element which switches  $e_3$  and  $e_6$ . Then  $w\Phi_0$  is orthogonal to  $e_1 + e_2 - 3e_6 + e_8$ . Let  $\Psi$  be the standard root subsystem generated by  $e_6 + e_7$  and  $-\frac{1}{2}(e_1 + \dots + e_8)$ ; then  $\Psi$  is of type  $A_2$ . However, no element in  $\Psi$  is orthogonal to  $e_1 + e_2 - 3e_6 + e_8$ , so  $w\Phi_0 \cap \Psi = \emptyset$ . Thus, like before, (52) fails to hold.  $\square$

**Lemma 7.8.** *Let  $\Phi$  be of type  $E_7$ , and let  $\Phi_0$  be an extremal root subsystem (necessarily of type  $E_6$ ). Then  $\Phi_0$  is semi-dense.*

*Proof.* We have as a model for the  $E_7$  root system those vectors in  $\mathbb{R}^8$  of the form  $e_i - e_j$  with  $i \neq j$  as well as all permutations of  $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, -\frac{1}{2}, -\frac{1}{2}, -\frac{1}{2}, -\frac{1}{2})$ . A base of simple roots is given by

$$\begin{aligned} & \{-e_2 + e_3, -e_3 + e_4, -e_4 + e_5, -e_5 + e_6, -e_6 + e_7, -e_7 + e_8, \\ & \quad \frac{1}{2}(e_1 + e_2 + e_3 + e_4 - e_5 - e_6 - e_7 - e_8)\}. \end{aligned}$$

Notice that all of  $\Phi$  is orthogonal to  $(1, \dots, 1)$ . Let  $\Phi_0$  be the standard root subsystem obtained by removing  $-e_7 + e_8$  from the base; then  $\Phi_0$  is of type  $E_6$  and exactly consists of those roots orthogonal to  $e_1 + e_8$ .

The Weyl group of  $E_7$  has order  $2^{10} \cdot 3^4 \cdot 5 \cdot 7$ , and that of  $E_6$  has order  $2^7 \cdot 3^4 \cdot 5$ . By the orbit-stabilizer theorem, we get that the  $W$ -orbit of  $e_1 + e_8$  is of size  $2^3 \cdot 7$ . These are exactly the  $2^2 \cdot 7$  vectors of the form  $e_i + e_j$  with  $i \neq j$ , as well as the  $2^2 \cdot 7$  vectors of the form  $\frac{1}{2}(e_1 + \dots + e_8) - (e_i + e_j)$  with  $i \neq j$ .

From here it is now easy to formulate a simple algorithm to verify (52). Each standard root subsystem can be found by choosing a subset of the base, and then seeing which roots lie in their span. Then for each such standard root subsystem  $\Psi$  and each vector  $v = w.(e_1 + e_8)$  in the  $W$ -orbit of  $e_1 + e_8$  described in the previous paragraph, we see how many roots are orthogonal to  $v$ . This tells us  $|\Psi \cap w\Phi_0|$ , and thus we can easily check (52). We carried out this algorithm in Sage [Ste+23] and found that in all cases (52) was satisfied.  $\square$

## 8. BOUNDING THE ELEMENTARY SPECTRAL INTEGRAL

The purpose of this section is to prove Proposition 6.2. Let  $r$  be the rank of  $G$ . We define coordinates on  $\mathfrak{a}_+^*$  by the map  $x \in \mathbb{R}_{>0}^r \mapsto \lambda(x) = x_1 \varpi_1 + \dots + x_r \varpi_r \in \mathfrak{a}_+^*$ . Applying this change of coordinates, the integral in (50) becomes

$$(54) \quad \int_{\mathbb{R}_{>0}^r} (1 + \|x\|)^{-N} \Theta(tH_0, \lambda(x))^2 \Theta(H, \lambda(x)) |c(\lambda(x))|^{-2} dx.$$

We shall bound this integral by dividing  $\mathbb{R}_{>0}^r$  into various subregions. As we shall mostly just work with reduced root systems in this section, we shall simplify notation by dropping the subscript  $\Phi_{\text{red}}$  everywhere, so that  $\Phi, \Phi_M, \Phi_M^+$ , etc. denote sets of reduced roots.

**8.1. Barycentric subdivision and corresponding estimates.** The first subdivision will be a barycentric one, whose subregions are determined by an ordering on the  $x_i$ . Let  $\sigma \in S_r$  be a permutation, and define  $T_\sigma$  to be the subset of  $\mathbb{R}_{>0}^r$  where  $x_{\sigma(1)} > \dots > x_{\sigma(r)}$ .

For  $1 \leq i \leq r$ , define  $\Phi_{\sigma,i}$  to be the set of roots orthogonal to  $\varpi_{\sigma(1)}, \dots, \varpi_{\sigma(i)}$ . In particular,  $\Phi_{\sigma,r} = \emptyset$ . We furthermore set  $\Phi_{\sigma,0} = \Phi$ . We observe that  $\Phi_{\sigma,i}$  is simply the standard sub-root system consisting of all roots expressible as linear combinations of simple roots in  $\Delta \setminus \{\alpha_{\sigma(1)}, \dots, \alpha_{\sigma(i)}\}$ .

**Lemma 8.1.** *Let  $1 \leq i \leq r$ . If  $\alpha \in \Phi_{\sigma,i-1} \setminus \Phi_{\sigma,i}$ , then  $|\langle \lambda(x), \alpha \rangle| \asymp x_{\sigma(i)}$  on  $T_\sigma$ .*

*Proof.* We may assume that  $\alpha \in \Phi^+$ . Let  $e_1, \dots, e_r$  be the standard basis for  $\mathbb{R}^r$ . If  $x \in T_\sigma$ , then  $x = x_{\sigma(1)}e_{\sigma(1)} + \dots + x_{\sigma(r)}e_{\sigma(r)}$  with  $x_{\sigma(1)} > \dots > x_{\sigma(r)}$ . We have  $\langle \alpha, \varpi_{\sigma(j)} \rangle = 0$  for

$1 \leq j \leq i-1$ , and  $\langle \alpha, \varpi_{\sigma(j)} \rangle > 0$  for  $i \leq j \leq r$ , so that

$$\begin{aligned}\langle \lambda(x), \alpha \rangle &= \langle x_{\sigma(1)} \varpi_{\sigma(1)} + \cdots + x_{\sigma(n-1)} \varpi_{\sigma(r)}, \alpha \rangle \\ &= x_{\sigma(i)} \langle \varpi_{\sigma(i)}, \alpha \rangle + \cdots + x_{\sigma(r)} \langle \varpi_{\sigma(r)}, \alpha \rangle \asymp x_{\sigma(i)},\end{aligned}$$

as required.  $\square$

We now use Lemma 8.1 to estimate the various terms appearing in (54) on  $T_\sigma$ .

**Lemma 8.2.** *Let  $\sigma \in S_r$ . Then, for  $x \in T_\sigma$  and  $\|H\| \ll t$ , we have*

$$(55) \quad \Theta(H, \lambda(x)) \ll \prod_{i=1}^r \min(t, x_{\sigma(i)}^{-1} + 1)^{|\Phi_{\sigma,i-1}^+| - |\Phi_{\sigma,i}^+|}.$$

Furthermore we have

$$(56) \quad \Theta(tH_0, \lambda(x)) \ll \sum_{w \in W} \prod_{i=1}^r \min(t, x_{\sigma(i)}^{-1} + 1)^{n_M(\sigma, i, w)},$$

where, for  $1 \leq i \leq r$ , we have put  $n_M(\sigma, i, w) = |(\Phi^+ \setminus \Phi_M^+) \cap w(\Phi_{\sigma,i-1} \setminus \Phi_{\sigma,i})|$ . Finally, we have

$$(57) \quad |c(\lambda(x))|^{-2} \ll (1 + x_{\sigma(1)})^d \prod_{i=1}^r x_{\sigma(i)}^{2|\Phi_{\sigma,i-1}^+| - 2|\Phi_{\sigma,i}^+|},$$

where  $d := \sum_{\alpha \in \Phi^+} m_\alpha + m_{2\alpha}$  is the sum of the multiplicities of all relative roots.

*Proof.* We begin with  $\Theta(H, \lambda(x))$ . We first apply the trivial estimate  $\alpha(H) \ll t$ , which gives

$$\Theta(H, \lambda(x)) \ll \sum_{w \in W} \prod_{\alpha \in \Phi^+} \min(t, |\langle w\lambda(x), \alpha \rangle|^{-1} + 1).$$

We next break the product over  $\Phi^+$  into the subsets  $\Phi^+ \cap w(\Phi_{\sigma,i-1} \setminus \Phi_{\sigma,i})$  and apply Lemma 8.1 to each subset, which gives

$$\begin{aligned}\prod_{\alpha \in \Phi^+} \min(t, |\langle w\lambda(x), \alpha \rangle|^{-1} + 1) &\asymp \prod_{i=1}^r \prod_{\alpha \in \Phi^+ \cap w(\Phi_{\sigma,i-1} \setminus \Phi_{\sigma,i})} \min(t, x_{\sigma(i)}^{-1} + 1) \\ &= \prod_{i=1}^r \min(t, x_{\sigma(i)}^{-1} + 1)^{|\Phi_{\sigma,i-1}^+| - |\Phi_{\sigma,i}^+|},\end{aligned}$$

yielding (55).

Next, we handle the  $c$ -function bound. When  $|\langle \lambda(x), \alpha \rangle|$  is small, by (19) we have that

$$|c_\alpha(\lambda(x))|^{-2} \ll |\langle \lambda(x), \alpha \rangle|^2.$$

When  $|\langle \lambda(x), \alpha \rangle|$  is big, by (20) we instead have that

$$|c_\alpha(\lambda(x))|^{-2} \ll |\langle \lambda(x), \alpha \rangle|^{m_\alpha + m_{2\alpha}} \ll (1 + x_{\sigma(1)})^{m_\alpha + m_{2\alpha}}.$$

We can combine these bounds and say that for all  $\lambda(x) \in T_\sigma$  we have

$$|c(\lambda(x))|^{-2} \ll (1 + x_{\sigma(1)})^d \prod_{\alpha \in \Phi^+} |\langle \lambda(x), \alpha \rangle|^2.$$

Applying the same argument used to prove (55) we find (57).

Finally, we apply a similar argument to  $\Theta(tH_0, \lambda)$ , with the added observation that  $\alpha(tH_0) = 0$  for  $\alpha \in \Phi_M^+$ . These roots therefore make no contribution to the formula for  $\Theta(tH_0, \lambda)$ , while for  $\alpha \in \Phi^+ \setminus \Phi_M^+$  we have  $\alpha(tH_0) \ll t$  as before. It follows that

$$\Theta(tH_0, \lambda(x)) \ll \sum_{w \in W} \prod_{\alpha \in \Phi^+ \setminus \Phi_M^+} \min(t, |\langle w\lambda(x), \alpha \rangle|^{-1} + 1).$$

As before, we partition  $\Phi^+ \setminus \Phi_M^+$  into the subsets  $(\Phi^+ \setminus \Phi_M^+) \cap w(\Phi_{\sigma, i-1} \setminus \Phi_{\sigma, i})$ . This gives

$$\begin{aligned} \prod_{\alpha \in \Phi^+ \setminus \Phi_M^+} \min(t, |\langle w\lambda(x), \alpha \rangle|^{-1} + 1) &\asymp \prod_{i=1}^r \prod_{\substack{\alpha \in (\Phi^+ \setminus \Phi_M^+) \\ \cap w(\Phi_{\sigma, i-1} \setminus \Phi_{\sigma, i})}} \min(t, x_{\sigma(i)}^{-1} + 1) \\ &= \prod_{i=1}^r \min(t, x_{\sigma(i)}^{-1} + 1)^{n_M(\sigma, i, w)}, \end{aligned}$$

yielding (56).  $\square$

**8.2. Refined subdivision and corresponding estimates.** We now refine the estimates in Lemma 8.2 based on which of the  $x_{\sigma(i)}$  are larger or smaller than  $t^{-1}$ . For  $0 \leq l \leq r$  we let

$$T_{\sigma, l} = \begin{cases} \{x \in T_\sigma : t^{-1} > x_{\sigma(1)}\}, & l = 0; \\ \{x \in T_\sigma : x_{\sigma(l)} > t^{-1} > x_{\sigma(l+1)}\}, & 1 \leq l \leq r-1; \\ \{x \in T_\sigma : x_{\sigma(r)} > t^{-1}\}, & l = r. \end{cases}$$

Note that  $l$  records how many  $x_{\sigma(i)}$ 's are bigger than  $t^{-1}$ .

It will be convenient to introduce the following notation.

**Definition 8.3.** For  $\sigma \in S_r$ ,  $w \in W$ , and  $1 \leq i \leq r$ , let

$$s_M(\sigma, i, w) = |\Phi_{\sigma, i-1}^+| - |\Phi_{\sigma, i}^+| - 2n_M(\sigma, i, w),$$

where  $n_M(\sigma, i, w)$  was defined in Lemma 8.2. For  $1 \leq i \leq r+1$  put

$$S_M(\sigma, i, w) = |\Phi_{\sigma, i}^+| - 2|(\Phi^+ \setminus \Phi_M^+) \cap w\Phi_{\sigma, i}|.$$

In particular, since  $\Phi_{\sigma, r} = \emptyset$ , we have  $S_M(\sigma, r+1, w) = 0$ , and since  $\Phi_{\sigma, 0} = \Phi$  we have  $S_M(\sigma, 1, w) = |\Phi^+| - 2|(\Phi^+ \setminus \Phi_M^+)| = |\Phi^+| - 2(|\Phi^+| - |\Phi_M^+|) = -|\Phi^+| + 2|\Phi_M^+|$ .

**Lemma 8.4.** Let  $\sigma \in S_r$ ,  $w \in W$ , and  $1 \leq l \leq r+1$ . There exists an  $N'$  such that for  $\lambda(x) \in T_{\sigma, l}$  and  $\|H\| \ll t$  we have

$$\Theta(tH_0, \lambda(x))^2 \Theta(H, \lambda(x)) |c(\lambda(x))|^{-2} \ll (1 + x_{\sigma(1)})^{N'} \left( \sum_{w \in W} t^{-S_M(\sigma, l, w)} \prod_{i=1}^l x_{\sigma(i)}^{s_M(\sigma, i, w)} \right),$$

where here, and throughout, a product from  $i = 1$  to  $l$  is assigned the value 1 when  $l = 0$ .

*Proof.* On  $T_{\sigma, l}$  we have  $x_{\sigma(i)} > t^{-1}$  for  $i \leq l$  and  $x_{\sigma(i)} < t^{-1}$  for  $i \geq l+1$ , and so

$$\begin{aligned} \min(t, x_{\sigma(i)}^{-1} + 1) &\asymp x_{\sigma(i)}^{-1} + 1 = x_{\sigma(i)}^{-1}(1 + x_{\sigma(i)}), \quad i \leq l, \\ \min(t, x_{\sigma(i)}^{-1} + 1) &= t \leq t(1 + x_{\sigma(i)}), \quad i \geq l+1. \end{aligned}$$

As a result, we deduce from the bound (55) for  $\Theta(H, \lambda)$  that

$$\Theta(H, \lambda(x)) \ll (1 + x_{\sigma(1)})^{|\Phi^+|} t^{|\Phi_{\sigma,l}^+|} \prod_{i=1}^l x_{\sigma(i)}^{-|\Phi_{\sigma,i-1}^+| + |\Phi_{\sigma,i}^+|}.$$

Likewise, the bound (56) for  $\Theta(tH_0, \lambda)$  becomes

$$\Theta(tH_0, \lambda(x)) \ll (1 + x_{\sigma(1)})^{|\Phi^+ \setminus \Phi_M^+|} \sum_{w \in W} t^{(|\Phi^+ \setminus \Phi_M^+|) \cap w\Phi_{\sigma,l}} \prod_{i=1}^l x_{\sigma(i)}^{-n_M(\sigma, i, w)}.$$

We simplify the bound (57) for  $|c(\lambda(x))|^{-2}$  to

$$|c(\lambda(x))|^{-2} \ll (1 + x_{\sigma(1)})^d t^{-2|\Phi_{\sigma,l}^+|} \prod_{i=1}^l x_{\sigma(i)}^{2|\Phi_{\sigma,i-1}^+| - 2|\Phi_{\sigma,i}^+|}.$$

Combining these estimates yields the lemma.  $\square$

For  $1 \leq l \leq r$ , we put

$$(58) \quad I_{M,\sigma,l,w}(t) = \int_{x_{\sigma(1)} > \dots > x_{\sigma(l)} > t^{-1}} (1 + x_{\sigma(1)})^{-N} \prod_{i=1}^l x_{\sigma(i)}^{s_M(\sigma, i, w)} dx_{\sigma(1)} \dots dx_{\sigma(l)}$$

and set  $I_{M,\sigma,0,w}(t) = 1$ .

**Corollary 8.5.** *In the above notation, and under the conditions of Lemma 8.4, we have, for  $N''$  much bigger than  $N'$ ,*

$$\int_{T_{\sigma,l}} (1 + x_{\sigma(1)})^{-N''} \Theta(tH_0, \lambda(x))^2 \Theta(H, \lambda(x)) |c(\lambda(x))|^{-2} dx \ll \sum_{w \in W} t^{-S_M(\sigma, l, w) + l - r} I_{M,\sigma,l,w}(t).$$

*Proof.* From Lemma 8.4 it follows that the left-hand side is bounded by

$$\sum_{w \in W} t^{-S_M(\sigma, l, w)} \int_{T_{\sigma,l}} (1 + x_{\sigma(1)})^{-N} \prod_{i=1}^l x_{\sigma(i)}^{s_M(\sigma, i, w)} dx_{\sigma(1)} \dots dx_{\sigma(r)}.$$

When  $l = r$  this gives the stated bound. For  $0 \leq l \leq r - 1$ , we perform the integrals over the variables which don't appear in the product, to find

$$\int_0^{t^{-1}} \int_0^{x_{\sigma(l+1)}} \dots \int_0^{x_{\sigma(r)}} dx_{\sigma(l+1)} \dots dx_{\sigma(r)} \leq \int_0^{t^{-1}} \dots \int_0^{t^{-1}} dx_{\sigma(l+1)} \dots dx_{\sigma(r)} = t^{l-r},$$

as desired.  $\square$

**8.3. Intersection cardinalities of root systems.** Before proceeding to estimate the integrals appearing in Corollary 8.5, we establish a few important properties of the quantities  $s_M(\sigma, i, w)$  and  $S_M(\sigma, i, w)$  appearing in Definition 8.3.

**Lemma 8.6.** *Let  $\sigma \in S_r$ ,  $w \in W$ , and  $\Phi_M$  a standard sub-root system. Then*

- (1) *for  $1 \leq i \leq r$  we have  $S_M(\sigma, i, w) + s_M(\sigma, i, w) = S_M(\sigma, i - 1, w)$ ;*
- (2) *for  $0 \leq i \leq r$  we have  $S_M(\sigma, i, w) = |\Phi_M \cap w\Phi_{\sigma,i}| - \frac{1}{2}|\Phi_{\sigma,i}|$ .*

*Proof.* For point (1) we have

$$\begin{aligned}
S_M(\sigma, i-1, w) - S_M(\sigma, i, w) &= |\Phi_{\sigma, i-1}^+| - 2|(\Phi^+ \setminus \Phi_M^+) \cap w\Phi_{\sigma, i-1}| \\
&\quad - (|\Phi_{\sigma, i}^+| - 2|(\Phi^+ \setminus \Phi_M^+) \cap w\Phi_{\sigma, i}|) \\
&= |\Phi_{\sigma, i-1}^+| - |\Phi_{\sigma, i}^+| + 2|(\Phi^+ \setminus \Phi_M^+) \cap w(\Phi_{\sigma, i-1} \setminus \Phi_{\sigma, i})| \\
&= s_M(\sigma, i, w).
\end{aligned}$$

For point (2) we calculate

$$\begin{aligned}
S_M(\sigma, i, w) &= |\Phi_{\sigma, i}^+| - 2|(\Phi^+ \setminus \Phi_M^+) \cap w\Phi_{\sigma, i}^+| \\
&= \frac{1}{2}|\Phi_{\sigma, i}| - |(\Phi \setminus \Phi_M) \cap w\Phi_{\sigma, i}| \\
&= \frac{1}{2}|\Phi_{\sigma, i}| - (|w\Phi_{\sigma, i}| - |\Phi_M \cap w\Phi_{\sigma, i}|),
\end{aligned}$$

which simplifies to the stated identity.  $\square$

**Corollary 8.7.** *Suppose that  $\Phi$  is of type  $A_n, B_n, C_n, D_n$ , or  $E_7$ . Let  $H_0$  be extremely singular with corresponding root system  $\Phi_M$ . Fix  $\sigma \in S_r$ ,  $w \in W$ , and  $0 \leq i \leq r$ . Then  $S_M(\sigma, i, w) + r \geq i$ .*

*Proof.* We apply Lemma 8.6(2), and then Lemma 7.5 and Lemma 7.8, with  $\Phi_0 = \Phi_M$  and  $\Psi = w\Phi_{\sigma, i}$ .  $\square$

**8.4. Integrating over the refined subdivision.** By contrast with the preceding steps, in which  $H_0$  was allowed to be arbitrary, the remainder of the argument will rely critically on the extremal hypothesis on the directing element  $H_0$ .

Recall the integral  $I_{M, \sigma, l, w}(t)$  defined in (58). When combined with Corollary 8.5, the following lemma completes the proof of Proposition 6.2.

**Lemma 8.8.** *Let  $G$  not be of type  $E_6, E_8, F_4$ , or  $G_2$ . Let  $\Phi_M$  be the root system corresponding to an extremely singular  $H_0$ . Fix  $\sigma \in S_r$ ,  $w \in W$ . For  $1 \leq l \leq r$  we let*

$$e_M(\sigma, l, w) = \#\{1 \leq i \leq l : S_M(\sigma, i, w) + r = i\}.$$

and put  $e_M(\sigma, 0, w) = 0$ . Then

$$t^{-S_M(\sigma, l, w)+l-r} I_{M, \sigma, l, w}(t) \ll (\log t)^{e_M(\sigma, l, w)},$$

the implied constant depending only on the integer  $N$  appearing in the definition of  $I_{M, \sigma, l, w}(t)$ .

*Proof.* For notational simplicity, we drop the dependence on  $w$ ,  $\sigma$ , and  $M$  in the notation  $I_{M, \sigma, l, w}(t)$ ,  $S_M(\sigma, l, w)$  and  $s_M(\sigma, i, w)$ , writing simply  $I_l(t)$ ,  $S(l)$  and  $s(i)$  for these quantities. In particular,  $S(r) = 0$ .

For  $l = 0$ , recall from the discussion preceding Corollary 8.5 that, by convention,  $I_0(t) = 1$ . We must therefore show that  $S(0) + r \geq 0$ , which follows from Corollary 8.7.

For  $1 \leq l \leq r$  we let

$$J_l(t) = \int_{x_{\sigma(1)} > \dots > x_{\sigma(l)} > t^{-1}} (1 + x_{\sigma(1)})^{-N} x_{\sigma(l)}^{S(l)+r-l} \prod_{i=1}^l x_{\sigma(i)}^{s(i)} dx_{\sigma(1)} \dots dx_{\sigma(l-1)}.$$

As was the case for  $I_{M,\sigma,l,w}(t)$ , we do not indicate the dependence on  $N$  in the notation, and we shall allow  $N$  to be taken sufficiently large. From the inequality  $(t^{-1})^{S(l)+r-l} < x_{\sigma(l-1)}^{S(l)+r-l}$ , it follows that  $t^{-S(l)+l-r} I_l(t) < J_l(t)$ . We shall bound  $J_l(t)$  by induction.

The base case is when  $l = 1$ , which we handle first. Since  $N$  can be taken sufficiently large, and  $S(1) + s(1) = S(0)$  from Lemma 8.6(1), we have

$$J_2(t) = \int_{t^{-1}}^{\infty} (1+x_{\sigma(1)})^{-N} x_{\sigma(1)}^{S(1)+s(1)+r-1} dx_{\sigma(1)} \ll \int_{t^{-1}}^1 x_{\sigma(1)}^{S(1)+s(1)+r-1} dx_{\sigma(1)} = \int_{t^{-1}}^1 x_{\sigma(1)}^{S(0)+r-1} dx_{\sigma(1)}.$$

From Corollary 8.7 the exponent satisfies  $S(0) + r - 1 \geq -1$ . If this exponent is  $\geq 0$  then we may extend the lower limit of integration to 0 and obtain a bound of 1. When the exponent is  $-1$ , the integral is bounded by  $\log t$ . This establishes the lemma when  $l = 1$ .

We now assume that  $2 \leq l \leq r$  and aim to show that

$$(59) \quad J_l(t) \ll (\log t)^{\delta_l} J_{l-1}(t), \quad \text{where} \quad \delta_l = \begin{cases} 0, & S(l-1) + r > l-1; \\ 1, & S(l-1) + r = l-1. \end{cases}$$

For this we consider the inner integral over  $x_{\sigma(l)}$  in  $J_l(t)$ . Using the recursive identity  $S(l) + s(l) = S(l-1)$  from Lemma 8.6(1) we have

$$\int_{t^{-1}}^{x_{\sigma(l-1)}} x_{\sigma(l)}^{S(l)+r-l+s(l)} dx_{\sigma(l)} = \int_{t^{-1}}^{x_{\sigma(l-1)}} x_{\sigma(l)}^{S(l-1)+r-l} dx_{\sigma(l-1)}.$$

From Corollary 8.7 the exponent satisfies  $S(l-1) + r - l \geq -1$ , with equality occurring precisely when  $\delta_l = 1$ . If  $\delta_l = 0$ , then this exponent is  $\geq 0$  and we may extend the lower limit of integration to 0 and obtain a bound of  $x_{\sigma(l-1)}^{S(l-1)+r-(l-1)}$ , yielding (59). If  $\delta_l = 1$ , then the inner integration over  $x_{\sigma(l-1)}$  now gives

$$\int_{t^{-1}}^{x_{\sigma(l-1)}} x_{\sigma(l)}^{-1} dx_{\sigma(l)} = \log x_{\sigma(l-1)} - \log(t^{-1}) \ll (1 + x_{\sigma(1)}) + \log t \ll (\log t)(1 + x_{\sigma(1)}).$$

The  $\log t$  term may be brought to the outside of the integral, and the  $1 + x_{\sigma(1)}$  may be absorbed, yielding

$$J_l(t) \ll (\log t) \int_{x_{\sigma(1)} > \dots > x_{\sigma(l-1)} > t^{-1}} (1 + x_{\sigma(1)})^{-N} \prod_{i=1}^{l-1} x_{\sigma(i)}^{s(i)} dx_{\sigma(1)} \dots dx_{\sigma(l-1)}.$$

Since, in the definition of  $J_{l-1}$ , the power of  $x_{\sigma(l-1)}$  in front of the product  $\prod_{i=1}^{l-1} x_{\sigma(i)}^{s(i)}$  is  $S(l-1) + r - (l-1) = 0$  when  $\delta_l = 1$ , the integral above is  $(\log t) J_{l-1}(t)$ , establishing (59) in this case as well.  $\square$

*Remark 8.9.* Though our proof techniques show that  $\text{vol}(e^H S_t \cap S_t) \leq (\log t)^k e^{\rho(2tH_0 - H)}$  for extremal  $H_0$  (or more generally for those  $H_0$  for which the roots vanishing on  $H_0$  forms a semi-dense root subsystem), our techniques allow one to more generally obtain a bound like  $\text{vol}(e^H S_t \cap S_t) \leq t^\ell e^{\rho(2tH_0 - H)}$  for general  $H_0$ .

## 9. BOUNDS FOR SPHERICAL FUNCTIONS

This section contains the proof of Theorems 2.3 and 2.4. Section 9.1 provides some illustration of these theorems, and explains why Theorem 2.4 should be sharp, and Section 9.2 discusses the relation between these theorems and previous work. Section 9.3 gives an

outline of the proof of Theorem 2.3, and the proof is carried out in Sections 9.4 to 9.6. We prove Theorem 2.4 in Section 9.7.

**9.1. Illustration of Theorems 2.3 and 2.4.** In this section, we make some remarks to help the reader understand the significance of Theorems 2.3 and 2.4. This includes explicating these theorems in the case of  $\mathrm{SL}_2(\mathbb{C})$ , deriving some simpler consequences of Theorem 2.3 for a general group, and explaining why Theorem 2.4 should be sharp based on stationary phase considerations. We also note that the bounds of Theorems 2.3 and 2.4 are equivalent in the case when  $G$  is a complex group and  $\lambda \in \mathfrak{a}^*$  is bounded, by Lemma 9.5 below.

**9.1.1. The case of  $\mathrm{SL}_2(\mathbb{C})$ .** When  $G = \mathrm{SL}_2(\mathbb{C})$ , we have the following explicit formula for the spherical function. If  $\alpha$  is the unique element of  $\Phi^+$ , we may identify  $\mathfrak{a}_{\mathbb{C}}^*$  with  $\mathbb{C}$  and  $\mathfrak{a}$  with  $\mathbb{R}$  using the basis elements  $\alpha$  and  $\alpha^\vee/2$  respectively. We then have

$$(60) \quad \varphi_\lambda(e^t) = \frac{\sin(\lambda t)}{\lambda \sinh t},$$

and we may compare Theorems 2.3 and 2.4 with the bounds that may be deduced from this formula. For simplicity, we shall assume that  $\lambda$  is real, and we assume without loss of generality that  $\lambda$  and  $t$  are non-negative. In this case, Theorem 2.3 states that

$$(61) \quad \varphi_\lambda(e^t) \ll (1 + \lambda)^a \min(t + 1, \lambda^{-1} + 1) e^{-t},$$

and Theorem 2.4 states

$$(62) \quad \varphi_\lambda(e^t) \ll e^{-t} \frac{t + 1}{t\lambda + 1}.$$

We note that these bounds are of the same strength if  $\lambda$  is bounded, as established in Lemma 9.5, and it may be checked that (62) is stronger than (61) as expected.

We shall show that (62) is sharp with complete uniformity in  $t$  and  $\lambda$ . To do this, we consider the cases  $t\lambda < 1$  and  $t\lambda \geq 1$  separately. When  $t\lambda < 1$  we have  $\sin(t\lambda) \asymp t\lambda$ , so that (60) implies that  $\varphi_\lambda(e^t) \asymp t/\sinh t$ . On the other hand, (62) gives

$$\varphi_\lambda(e^t) \ll (t + 1)e^{-t} \asymp t/\sinh t,$$

which shows that (62) is optimal in this range.

When  $t\lambda \geq 1$  the best bound one can give for  $\sin(\lambda t)$  is  $|\sin(\lambda t)| \leq 1$ , and so by (60) the best bound one can give for  $\varphi_\lambda(e^t)$  is  $\varphi_\lambda(e^t) \ll 1/(\lambda \sinh t)$ . In comparison, (62) becomes

$$\varphi_\lambda(e^t) \ll e^{-t} \frac{t + 1}{t\lambda},$$

and this is equivalent to  $\varphi_\lambda(e^t) \ll 1/(\lambda \sinh t)$  as one sees by applying the asymptotic  $1/\sinh t \asymp e^{-t}(t + 1)/t$ .

The bound (62) reflects the fact that the decay of  $\varphi_0(e^t) = t/\sinh(t) \asymp te^{-t}$  is slower than that of  $\varphi_\lambda$  for  $\lambda \neq 0$ , and moreover that this transition starts to be observed at the point  $e^t$  when  $\lambda \ll 1/t$ . Theorem 2.4 can be viewed as an extension of this bound to a general group, where the function  $(t + 1)/(t\lambda + 1)$  is replaced by a product of similar expressions over  $\Phi_{\mathrm{red}}^+$ .

9.1.2. *Consequences on a general group.* We next derive some simpler corollaries of Theorem 2.3, both because of their intrinsic interest, and to facilitate the comparison between Theorem 2.3 and previous bounds for the spherical function in the next section. First, we make the observation that when  $\lambda$  is at distance  $\gg 1$  from the singular set, Theorem 2.3 simply gives

$$(63) \quad \varphi_\lambda(e^H) \ll (1 + \|\lambda\|)^a \max_{w \in W} e^{-(\rho + w\Im\lambda)(H)}.$$

For general  $\lambda$ , we may apply the bound  $f_\alpha(H, \lambda) \leq |\alpha(H)| + 1$  to deduce

$$(64) \quad \varphi_\lambda(e^H) \ll (1 + \|\lambda\|)^a \max_{w \in W} e^{-(\rho + w\Im\lambda)(H)} \prod_{\alpha \in \Phi_{\text{red}}^+} (|\alpha(H)| + 1).$$

Theorem 2.3 may be viewed as giving an interpolation between these two bounds, based on how singular  $\lambda$  is. We may also apply the bound  $\alpha(H) \ll \|H\|$  to obtain

$$f_\alpha(H, w\lambda) \ll \min(\|H\| + 1, |\langle w\lambda, \alpha \rangle|^{-1} + 1),$$

which gives

$$\Theta(H, \lambda) \ll \prod_{\alpha \in \Phi_{\text{red}}^+} \min(\|H\| + 1, |\langle \lambda, \alpha \rangle|^{-1} + 1)$$

and therefore

$$(65) \quad \varphi_\lambda(e^H) \ll (1 + \|\lambda\|)^a \max_{w \in W} e^{-(\rho + w\Im\lambda)(H)} \prod_{\alpha \in \Phi_{\text{red}}^+} \min(\|H\| + 1, |\langle \lambda, \alpha \rangle|^{-1} + 1).$$

Finally, we have the following.

**Corollary 9.1.** *Let  $\eta$  denote the maximum size of the sets  $\Phi_{\text{red}, M}^+$ , where  $M$  runs over the maximal standard Levi subgroups of  $G$ . We then have*

$$\varphi_\lambda(e^H) \ll (1 + \|\lambda\|)^a (1 + \|H\|)^\eta \min(\|H\| + 1, \|\lambda\|^{-1} + 1)^{|\Phi_{\text{red}}^+| - \eta} \max_{w \in W} e^{-(\rho + w\Im\lambda)(H)}$$

*Proof.* We shall derive this from (65), by showing that

$$\prod_{\alpha \in \Phi_{\text{red}}^+} \min(\|H\| + 1, |\langle \lambda, \alpha \rangle|^{-1} + 1) \ll (1 + \|H\|)^\eta \min(\|H\| + 1, \|\lambda\|^{-1} + 1)^{|\Phi_{\text{red}}^+| - \eta}$$

for any  $\lambda \in \mathfrak{a}_\mathbb{C}^*$ . We are free to multiply  $\lambda$  by  $i$ , and so we may assume that  $\|\Re\lambda\| \gg \|\lambda\|$ . We may also assume that  $\Re\lambda \in \bar{\mathfrak{a}}_+^*$ . It follows that there is some maximal standard Levi subgroup  $M$  such that  $|\langle \lambda, \alpha \rangle| \gg \|\lambda\|$  for all  $\alpha \in \Phi_{\text{red}}^+ \setminus \Phi_{\text{red}, M}^+$ , and therefore that

$$\min(\|H\| + 1, |\langle \lambda, \alpha \rangle|^{-1} + 1) \ll \min(\|H\| + 1, \|\lambda\|^{-1} + 1)$$

for  $\alpha \in \Phi_{\text{red}}^+ \setminus \Phi_{\text{red}, M}^+$ . This implies

$$\begin{aligned} \prod_{\alpha \in \Phi_{\text{red}}^+} \min(\|H\| + 1, |\langle \lambda, \alpha \rangle|^{-1} + 1) &\ll \prod_{\alpha \in \Phi_{\text{red}, M}^+} (\|H\| + 1) \prod_{\alpha \in \Phi_{\text{red}}^+ \setminus \Phi_{\text{red}, M}^+} \min(\|H\| + 1, \|\lambda\|^{-1} + 1) \\ &= (1 + \|H\|)^{|\Phi_{\text{red}, M}^+|} \min(\|H\| + 1, \|\lambda\|^{-1} + 1)^{|\Phi_{\text{red}}^+ \setminus \Phi_{\text{red}, M}^+|} \\ &\leq (1 + \|H\|)^\eta \min(\|H\| + 1, \|\lambda\|^{-1} + 1)^{|\Phi_{\text{red}}^+| - \eta} \end{aligned}$$

as required.  $\square$

9.1.3. *Theorem 2.4 and stationary phase.* We now explain why the bound of Theorem 2.4 is to be expected, and should be sharp, from the viewpoint of stationary phase applied to the integral (5). The critical set of the phase function appearing in this integral is understood, see e.g. Section 1 of [DKV79]. In particular, when  $X = H \in \mathfrak{a}$ , the integral in (5) reduces to one over  $K/Z_K(\mathfrak{a})$ , and [DKV79] shows that when  $H$  and  $\lambda$  are both regular the critical points occur at  $W$  and are non-degenerate. Moreover, for each  $w \in W$ , there is a fixed basis for the tangent space  $T_w(K/Z_K(\mathfrak{a}))$  in which the Hessian is diagonal with eigenvalues  $\alpha(H)\langle\lambda, w\alpha\rangle$ , each occurring with multiplicity 2 (Prop. 1.4 of [DKV79]). (The fact that the multiplicities are 2 is due to  $G$  being complex.) By stationary phase, the strongest bound one can give for the contribution of the integral near  $w$  to (5) is therefore  $(1 + |\alpha(H)\langle\lambda, w\alpha\rangle|)^{-1}$ , and summing these over  $W$  gives the bound of Theorem 2.4.

**9.2. Relation with previous work.** We now discuss the relation between Theorem 2.3 and existing bounds for the spherical function. Compared with previous results, Theorem 2.3 has strong dependence on  $H$ , but weak dependence on  $\lambda$ , at least when  $\lambda$  is large. This is convenient for us, on both counts. It is crucial for the applications in this paper that our bounds have good dependence on  $H$ . On the other hand, the fact that the bound of Theorem 2.3 is polynomially growing in  $\lambda$  does not matter for us, as in our applications there are always other factors present that decay rapidly in  $\lambda$ .

On a general group, the problem of giving bounds for the spherical function that are uniform in  $H$  and  $\lambda$  is a difficult one, and far from being completely solved. If we restrict our attention to bounds that hold for all  $H$  and  $\lambda$ , with  $\lambda$  varying over  $\mathfrak{a}^*$  or  $\mathfrak{a}_C^*$ , the only previous results we are aware of are obtained by reducing directly to the case of  $\lambda = 0$ , and as a result do not capture most of the interaction between  $H$  and  $\lambda$ .

To describe these bounds, it follows from Kostant's convexity theorem that

$$|\varphi_\lambda(e^H)| \leq \varphi_0(e^H) \max_{w \in W} e^{-w \Im \lambda(H)}$$

for any  $\lambda \in \mathfrak{a}_C^*$ , see for instance Prop 4.6.1 of [GV88]. Combined with the Harish-Chandra bound for  $\varphi_0$  (Theorem 4.6.4 of [GV88]), which states that

$$(66) \quad \varphi_0(e^H) \ll (1 + \|H\|)^{|\Phi_{\text{red}}^+|} e^{-\rho(H)},$$

we obtain

$$(67) \quad \varphi_\lambda(e^H) \ll (1 + \|H\|)^{|\Phi_{\text{red}}^+|} \max_{w \in W} e^{-(\rho + w \Im \lambda)(H)}.$$

The bound (66) was strengthened by Anker [Ank87] to

$$(68) \quad \varphi_0(e^H) \asymp \prod_{\alpha \in \Phi_{\text{red}}^+} (1 + |\alpha(H)|) e^{-\rho(H)},$$

which gives the corresponding improvement in (67) to

$$(69) \quad \varphi_\lambda(e^H) \ll \prod_{\alpha \in \Phi_{\text{red}}^+} (1 + |\alpha(H)|) \max_{w \in W} e^{-(\rho + w \Im \lambda)(H)}.$$

We note that the upper bound for  $\varphi_0$  provided by Theorem 2.3 is the same as the upper bound in (68), so that the theorem is sharp in this case. Moreover, (69) is equivalent to the corollary (64) of Theorem 2.3, if we restrict  $\lambda$  to a bounded set in  $\mathfrak{a}^*(\kappa)$ .

There are also previous bounds for  $\varphi_\lambda(e^H)$  proved using the Harish-Chandra expansion. Such bounds always blow up along the walls of  $\bar{\mathfrak{a}}_+$ , and so they are generally stated for  $H$  in a set of the form  $H_0 + \mathfrak{a}_+$  for a fixed  $H_0 \in \mathfrak{a}_+$ . To our knowledge, the strongest bound of this type is proved in work in progress of the second author, and we state it in a simplified form here. Roughly speaking, it applies a complexification argument to the Harish-Chandra series expansion of  $\varphi_\lambda$ , similar to the one used in this paper with the expansion of Gangolli–Varadarajan. As a result, it has better dependence on  $\lambda$ , but worse dependence on  $H$  (and does not hold near the walls as mentioned above). To state it, for  $\alpha \in \Phi_{\text{red}}$ , define the function  $\Theta_\alpha^0$  on  $\mathfrak{a} \times \mathfrak{a}_+^*$  by

$$\Theta_\alpha^0(H, \lambda) = \begin{cases} \|H\| & |\langle \lambda, \alpha \rangle| \leq (1 + \|H\|)^{-1}, \\ |\langle \lambda, \alpha \rangle|^{-1} & (1 + \|H\|)^{-1} < |\langle \lambda, \alpha \rangle| \leq 1, \\ |\langle \lambda, \alpha \rangle|^{-(m_\alpha + m_{2\alpha})/2} & 1 < |\langle \lambda, \alpha \rangle|. \end{cases}$$

We also define

$$\Theta^0(H, \lambda) = \prod_{\alpha \in \Phi_{\text{red}}^+} \Theta_\alpha^0(H, \lambda).$$

Let  $H_0 \in \mathfrak{a}_+$  and  $R > 0$  be given. The bound then states that

$$\varphi_\lambda(e^H) \ll_{H_0, R} \Theta^0(H, \lambda) \|H\|^{\eta(R)} \max_{w \in W} e^{-(\rho + w\Im\lambda)(H)}$$

for  $\lambda \in \mathfrak{a}^*(R)$  and  $H \in H_0 + \mathfrak{a}_+$ , where  $\eta(R) \geq 0$  is a constant that is equal to 0 if  $R$  is sufficiently small depending on  $G$ . There is also a version of the result that bounds the derivatives of  $\varphi_\lambda(e^H)$  with respect to  $H$  and  $\lambda$ .

In [NPP14], the Harish-Chandra expansion was used to prove an asymptotic for  $\varphi_\lambda$  similar to (68) for any  $\lambda$ , again subject to the condition that  $H$  lies in  $H_0 + \mathfrak{a}_+$  for a fixed  $H_0 \in \mathfrak{a}_+$ , and with error terms that are not uniform in  $\lambda$ . We have not checked how this asymptotic compares with Theorem 2.3. We also mention Cor. 4.5.5 of [GV88], which implies a uniform estimate for  $\varphi_\lambda(e^H)$  when  $H \in H_0 + \mathfrak{a}_+$  and  $\lambda \in \lambda_0 + \mathfrak{a}_+^*$  for fixed  $\lambda_0 \in \mathfrak{a}_+^*$ .

Finally, we assume that  $G$  is complex, and discuss the strength of Theorem 2.4, as well as its relation to existing bounds for the spherical function in this case.

Bounds for  $\varphi_\lambda$  and  $\varphi_\lambda^E$  in the complex case were studied by Barlet–Clerc [BC86], Clerc [Cle87], and Cowling–Nevo [CN01], and may be deduced from the results of Duistermaat–Kolk–Varadarajan [DKV79]. Theorem 2.4 strengthens these results, subject to the condition that  $\lambda$  is real. However, we note that the papers [BC86; Cle87; CN01] also obtain results for complex  $\lambda$ , and [CN01] also bounds certain derivatives of  $\varphi_\lambda(e^H)$  in the  $H$ -variable.

**9.3. Outline of proof of Theorem 2.3.** We now give an outline of the proof of Theorem 2.3. We first discuss the case when  $G = \text{SL}_2(\mathbb{R})$ , before describing the modifications required in the general case. As in the case of  $\text{SL}_2(\mathbb{C})$  discussed above, we shall identify  $\mathfrak{a}$  and  $\mathfrak{a}_+^*$  with  $\mathbb{R}$  and  $\mathbb{C}$  using the basis elements  $\alpha^\vee/2$  and  $\alpha$  respectively, where  $\alpha \in \Phi^+$  is the unique positive root. Thus we wish to show that there exist  $a, \kappa > 0$  such that

$$(70) \quad |\varphi_\lambda(e^t)| \ll (1 + |\lambda|)^a e^{-t/2 + |\Im\lambda|t} \min(1 + t, 1 + |\lambda|^{-1})$$

for  $t \geq 0$  and  $|\Im\lambda| < \kappa$ .

First, we note that  $|\varphi_\lambda(e^t)| \leq e^{|\Im\lambda|t} \varphi_0(e^t)$ , which follows by taking absolute values in the Harish-Chandra integral formula and using the fact that  $H(ke^t) \leq t$  for all  $k \in \mathrm{SO}(2)$ . This implies (70) in the range where  $t < 1$  (or  $t < R$  for any fixed  $R$ ).

Now suppose  $t \geq 1$ . Let  $\theta(t, \lambda)$  be defined as

$$(71) \quad \theta(t, \lambda) := c(\lambda)e^{i\lambda t} + c(-\lambda)e^{-i\lambda t}.$$

The asymptotic formula of Gangolli–Varadarajan, recalled in Proposition 3.2, states that there is  $\kappa > 0$  such that

$$(72) \quad |\varphi_\lambda(e^t) - e^{-t/2}\theta(t, \lambda)| \ll (1+t)^a(1+|\lambda|)^a e^{-(1/2+\varepsilon)t}$$

for  $|\Im\lambda| < \kappa$ , where  $a, \varepsilon > 0$  are constants (see Section 5.1 of [GV88] for an illustrative discussion of the rank one case). We may take these to be the constants  $a$  and  $\kappa$  in our bound (70). As the right hand side of (72) is less than the right hand side of (70), it suffices to show that

$$(73) \quad |\theta(t, \lambda)| \ll (1+|\lambda|)^a e^{|\Im\lambda|t} \min(1+t, 1+|\lambda|^{-1}),$$

for  $t \geq 1$  and  $|\Im\lambda| < \kappa$ . In fact, we will show the stronger bound

$$(74) \quad |\theta(t, \lambda)| \ll e^{|\Im\lambda|t} \min(1+t, 1+|\lambda|^{-1}).$$

In the case of  $\mathrm{SL}_2(\mathbb{R})$ , the explicit nature of  $\theta(t, \lambda)$  means that this can be exhibited quite directly, and we do this first before describing the argument that must be used in the case of a general group. We first suppose that  $|\lambda| \geq 1$ . If we assume that  $\kappa < 1$ , we have the bound  $|c(\lambda)| \ll_\kappa 1$  for  $|\Im\lambda| < \kappa$ . We thus straight away get

$$|\theta(t, \lambda)| \ll e^{|\Im\lambda|t} \ll e^{|\Im\lambda|t} \min(1+|\lambda|^{-1}, 1+t),$$

as required.

We next suppose that  $|\lambda| < 1$ . The  $c$ -function is a meromorphic function with a simple pole at  $\lambda = 0$ . Despite this singularity,  $\theta(t, \lambda)$  can be extended holomorphically to  $\lambda = 0$ . If we write  $c(\lambda) = a_{-1}\lambda^{-1} + g(\lambda)$  where  $g(\lambda)$  is an entire function, we have

$$c(\lambda)e^{i\lambda t} + c(-\lambda)e^{-i\lambda t} = a_{-1} \frac{e^{i\lambda t} - e^{-i\lambda t}}{\lambda} + g(\lambda)e^{i\lambda t} + g(-\lambda)e^{-i\lambda t}.$$

Because  $|\lambda| < 1$ , the last two terms above can be bounded by  $Ce^{|\Im\lambda|t}$  for some  $C$ . The first term can be bounded in two different ways, as

$$\left| \frac{e^{i\lambda t} - e^{-i\lambda t}}{\lambda} \right| \leq 2|\lambda|^{-1}e^{|\Im\lambda|t} \quad \text{and} \quad \left| \frac{e^{i\lambda t} - e^{-i\lambda t}}{\lambda} \right| \leq 2te^{|\Im\lambda|t},$$

where the first bound is elementary and the second follows from applying the mean value theorem to  $f(x) = e^{i\lambda x} - e^{-i\lambda x}$  on the interval  $[0, t]$ . We thus get the bound

$$|\theta(t, \lambda)| \ll e^{|\Im\lambda|t} \min(1+|\lambda|^{-1}, 1+t),$$

as required. This completes the proof of Theorem 2.3 in this case.

On a general group, it will be convenient to use a different method, based on the maximum modulus principle, to establish (74). To describe it, we first observe that if  $\kappa < 1$ , we have

$$c(\lambda) \ll_\kappa 1 + |\lambda|^{-1}$$

when  $|\Im \lambda| < \kappa$ , as follows from (20). This implies that

$$(75) \quad \theta(t, \lambda) \ll (1 + |\lambda|^{-1})e^{|\Im \lambda|t},$$

which establishes (74) when  $|\lambda| \geq t^{-1}$ . We therefore assume that  $|\lambda| < t^{-1}$ . For these  $\lambda$ , we shall prove (74) by applying the maximum modulus principle, on the disk of radius  $2t^{-1}$  centered at  $\lambda$ . This gives

$$(76) \quad \theta(t, \lambda) \leq \max_{\alpha} \theta(t, \lambda + 2t^{-1}e^{i\alpha}).$$

As discussed above, we are free to assume that  $t \geq R$  for any fixed  $R$ , and we choose  $R$  large enough that  $3R^{-1} < \kappa$ . This implies that the contour  $\lambda + 2t^{-1}e^{i\alpha}$  is contained in the strip  $|\Im z| < \kappa$ , and lies outside the disk  $|z| < t^{-1}$ . We may therefore combine (75) and (76) to obtain

$$\begin{aligned} \theta(t, \lambda) &\ll \max_{\alpha} (1 + |\lambda + 2t^{-1}e^{i\alpha}|^{-1})e^{|\Im(\lambda + 2t^{-1}e^{i\alpha})|t} \\ &\leq 2e^{|\Im \lambda|t} \max_{\alpha} (1 + |\lambda + 2t^{-1}e^{i\alpha}|^{-1}), \end{aligned}$$

and moreover, because  $|\lambda + 2t^{-1}e^{i\alpha}| \geq t^{-1}$ , this gives

$$\theta(t, \lambda) \ll (1 + t)e^{|\Im \lambda|t}$$

as required.

We finish this section by describing the modifications that must be made to this approach in the case of a general group, and we now let  $G$  be general again. We shall apply the asymptotic of Gangolli–Varadarajan along maximal Levi subgroups of  $G$ . To do this, for each maximal Levi  $L$  we define a cone  $\mathcal{C}_L$  in  $\mathfrak{a}$  that is adapted to the centralizer of  $L$  in  $\mathfrak{a}$ , and such that  $\bar{\mathfrak{a}}_+ \subset B_R(0) \cup \bigcup_L \mathcal{C}_L$ . As before, the case of  $H \in B_R(0)$  may be handled easily, so we fix  $L$  and bound  $\varphi_{\lambda}(e^H)$  for  $H \in \mathcal{C}_L$ . If we define

$$\theta_L(H, \lambda) = \sum_{w \in W_L \setminus W} c^L(w\lambda) \varphi_{w\lambda}^L(e^H),$$

then the asymptotic of Gangolli–Varadarajan reduces the problem to proving that

$$(77) \quad \theta_L(H, \lambda) \ll (1 + \|\lambda\|)^a \Theta(H, \lambda) \max_{w \in W} e^{-(\rho_L + w\Im \lambda)(H)}.$$

For this purpose, let  $\kappa > 0$  and let

$$(78) \quad \mathfrak{a}^*(\kappa) = \{\lambda \in \mathfrak{a}_{\mathbb{C}}^* : \|\Im \lambda\| < \kappa\}$$

denote the tube of radius  $\kappa$  about  $\mathfrak{a}^* \subset \mathfrak{a}_{\mathbb{C}}^*$ . For a constant  $C > 0$ , we define

$$(79) \quad \mathfrak{a}_{\text{reg}}^*(\kappa, C) = \{\lambda \in \mathfrak{a}^*(\kappa) : |\langle \alpha, \lambda \rangle| \geq C \ \forall \alpha \in \Phi_{\text{red}}\}$$

and  $\mathfrak{a}_{\text{sing}}^*(\kappa, C) = \mathfrak{a}^*(\kappa) \setminus \mathfrak{a}_{\text{reg}}^*(\kappa, C)$ . We may show (77) when  $\lambda \notin \mathfrak{a}_{\text{reg}}^*(\kappa, \|H\|^{-1})$  by applying Theorem 2.3 inductively to  $L$ , together with standard bounds for  $c^L$ . We then deduce (77) for  $\lambda \in \mathfrak{a}_{\text{reg}}^*(\kappa, \|H\|^{-1})$  from this using the maximum principle. We do this by choosing a contour around  $\lambda$  of the form  $\lambda + \nu e^{i\alpha}$ , for some  $\nu \in \mathfrak{a}_{\mathbb{C}}^*$  depending on  $\lambda$ , such that

$$\max_{\alpha} \Theta(H, \lambda + \nu e^{i\alpha}) \max_{w \in W} e^{-(\rho_L + w\Im(\lambda + \nu e^{i\alpha}))(H)} \ll \Theta(H, \lambda) \max_{w \in W} e^{-(\rho_L + w\Im \lambda)(H)}.$$

We require  $\nu$  to satisfy  $\|\nu\| \ll \|H\|^{-1}$ , so that the exponential terms are roughly constant, and also that  $\lambda + \nu e^{i\alpha}$  is at distance  $\gg \|H\|^{-1}$  from all root hyperplanes. In Lemma 9.2, we

show that we may do this by choosing  $\nu = C\|H\|^{-1}\rho$ , where  $C \asymp 1$  depends on  $\lambda$ . We take  $\rho$  as the direction of  $\nu$  because it is not orthogonal to any root, and we need to choose  $C$  depending on  $\lambda$  because we now have multiple root hyperplanes to avoid, and a choice of  $C$  that avoids one hyperplane may end up intersecting another.

**9.4. Preliminaries.** We recall the notation  $\mathfrak{a}^*(\kappa)$  and  $\mathfrak{a}_{\text{reg}}^*(\kappa, C)$  from (78) and (79).

**Lemma 9.2.** *Let  $0 < \kappa' < \kappa$  be given. Then there exists  $\sigma > 1$  depending only on  $G$ , and  $D > 0$  depending on  $\kappa, \kappa'$ , and  $G$  such that the following holds. For each  $s < D$  and  $\lambda \in \mathfrak{a}^*(\kappa')$ , there exists  $C = C(s, \lambda) \in [1, \sigma]$  such that we have  $\lambda + z\rho \in \mathfrak{a}_{\text{reg}}^*(\kappa, s)$  for  $|z| = Cs$ .*

*Proof.* Let  $\tau$  be a real number satisfying  $\tau|\langle \rho, \alpha \rangle| > 2$  and  $\tau > |\langle \rho, \alpha \rangle| + 1$  for all  $\alpha \in \Phi_{\text{red}}$  (so in particular  $\tau > 1$ ). We claim that we can take  $\sigma = \tau^{2|\Phi_{\text{red}}^+|+2}$ .

To prove this, let  $0 < \kappa' < \kappa$ , and let  $\lambda \in \mathfrak{a}^*(\kappa')$  be given. It is clear that if  $C \in [1, \sigma]$  and  $|z| = Cs$  for  $s$  sufficiently small depending on  $\kappa$  and  $\kappa'$ , then  $\lambda + z\rho \in \mathfrak{a}^*(\kappa)$ . We must therefore show that there is a choice of  $C \in [1, \sigma]$  such that  $|\langle \lambda + z\rho, \alpha \rangle| \geq s$  for all  $|z| = Cs$  and  $\alpha \in \Phi_{\text{red}}$ .

Consider the disjoint intervals  $(\tau^{2k}s, \tau^{2k+2}s)$  for  $0 \leq k \leq |\Phi_{\text{red}}^+|$ . Because there are at most  $|\Phi_{\text{red}}^+|$  numbers of the form  $|\langle \lambda, \alpha \rangle|$ , one of these intervals does not contain  $|\langle \lambda, \alpha \rangle|$  for any  $\alpha \in \Phi_{\text{red}}$ . We let  $C = \tau^{2k+1}$  where  $k$  is index of one of the empty intervals. For each  $\alpha$  our choice of  $k$  implies that either  $|\langle \lambda, \alpha \rangle| \leq \tau^{2k}s$  or  $|\langle \lambda, \alpha \rangle| \geq \tau^{2k+2}s$ . In the first case we have

$$\begin{aligned} |\langle \lambda + z\rho, \alpha \rangle| &\geq |z||\langle \rho, \alpha \rangle| - |\langle \lambda, \alpha \rangle| \\ &\geq \tau^{2k+1}s|\langle \rho, \alpha \rangle| - \tau^{2k}s \\ &= \tau^{2k}(\tau|\langle \rho, \alpha \rangle| - 1)s \\ &> s, \end{aligned}$$

while in the second case a similar computation gives

$$|\langle \lambda + z\rho, \alpha \rangle| \geq \tau^{2k+1}s(\tau - |\langle \rho, \alpha \rangle|) > s,$$

as required.  $\square$

**9.5. Proof of Theorem 2.3: Reduction step.** We will prove Theorem 2.3 by induction on the semisimple real rank of  $G$ . The base case, when the rank is zero, is trivial, since the spherical functions in this case are simply complex exponentials. We may therefore assume that Theorem 2.3 holds for all proper Levi subgroups of  $G$ .

Let  $B_R(0)$  denote the ball of radius  $R$  around the origin in  $\mathfrak{a}$ . As in the rank one case discussed above, Theorem 2.3 holds when  $H \in B_R(0)$ , for any fixed  $R$ ; this follows from the bound  $|\varphi_\lambda(e^H)| \leq \max_{w \in W} e^{-w \Im \lambda(H)} \varphi_0(e^H)$ , which in turn follows from e.g. Prop. 4.6.1 of [GV88]. For the other  $H$ , we shall cover  $\bar{\mathfrak{a}}_+ \setminus B_R(0)$  with cones adapted to the centers of the maximal Levi subgroups of  $G$ .

Let  $\Sigma = \Delta \setminus \{\beta\}$  for some  $\beta \in \Delta$ . Let  $L_\Sigma$  be the Levi associated to  $\Sigma$ , which is the centralizer of the subspace  $\mathfrak{a}_\Sigma = \{H \in \mathfrak{a} : \alpha(H) = 0, \alpha \in \Sigma\} = \mathbb{R}\beta^\vee$ . Let  $\Phi_\Sigma$  be the root system of  $L_\Sigma$ , which is given by  $\Phi_\Sigma = \Phi \cap \sum_{\alpha \in \Sigma} \mathbb{R}\alpha$ . We let  $\mathcal{C}_\Sigma \subset \mathfrak{a}$  be the cone consisting of those  $H$  satisfying

$$(80) \quad \alpha(H) > c\|H\| \text{ for } \alpha \in \Phi^+ \setminus \Phi_\Sigma^+, \quad \|H\| > R.$$

Roughly speaking, this is a cone around  $\mathfrak{a}_\Sigma \cap \bar{\mathfrak{a}}_+$ . If  $c$  is chosen small enough, the  $\mathcal{C}_\Sigma$  for various  $\Sigma$ , together with  $B_R(0)$ , cover  $\bar{\mathfrak{a}}_+$ . Thus, from now on assume that we have fixed such a sufficiently small  $c$ .

We fix a  $\Sigma$ , and denote  $L_\Sigma$  simply by  $L$ . We prefer to write  $\mathcal{C}_L$  for  $\mathcal{C}_\Sigma$  and  $\Phi_L$  for  $\Phi_\Sigma$ . We shall prove Theorem 2.3 for  $H \in \bar{\mathfrak{a}}_+ \cap \mathcal{C}_L$ , using our hypothesis that it holds for  $L$ , together with the following two ingredients. The first is the asymptotic expansion of  $\varphi_\lambda$  along  $L$  given in Theorem 5.9.3 of [GV88].

**Proposition 9.3.** *Let  $L$  be a standard Levi subgroup of  $G$  and put*

$$\theta_L(H, \lambda) = \sum_{w \in W_L \setminus W} c^L(w\lambda) \varphi_{w\lambda}^L(e^H).$$

*Then there is  $a, \kappa > 0$ , depending only on the constant  $c$  appearing in (80), such that for all  $\lambda \in \mathfrak{a}^*(\kappa)$  and  $H \in \mathcal{C}_L$ , we have*

$$(81) \quad \varphi_\lambda(e^H) = e^{-\rho^L(H)} \theta_L(H, \lambda) + O((1 + \|\lambda\|)^a e^{-\rho(H) - \kappa\|H\|}).$$

*Proof.* We note that  $\theta_L(H, \lambda)$  is initially defined only for  $\lambda$  regular, but it is shown in Prop. 5.8.2 of [GV88] that it extends holomorphically to  $\mathfrak{a}^*(\kappa)$  if  $\kappa$  is sufficiently small. (This result is stated for a vector-valued function denoted  $\Theta(\lambda, m)$  in [GV88], but  $\theta_L(H, \lambda)$  is equal to the first coordinate of  $\Theta(\lambda, e^H)$ .) This may also be shown in a more elementary way by proving that the poles of  $c^L$  along the root hyperplanes cancel in the sum.

For any  $\zeta > 0$ , Gangolli–Varadarajan [GV88] (5.9.4) define a set  $A^+(H_0 : \zeta) \subset A$ , which in our case satisfies

$$A^+(H_0 : \zeta) = \exp(\{H \in \bar{\mathfrak{a}}_+ : \alpha(H) > \zeta\|H\| \text{ for } \alpha \in \Phi^+ \setminus \Phi_L^+\}).$$

If we choose  $\zeta = c$ , then we have  $\exp(\mathcal{C}_L \cap \bar{\mathfrak{a}}_+) \subset A^+(H_0 : \zeta)$ . In Theorem 5.9.3(b) of [GV88], the authors prove an asymptotic for  $\varphi_\lambda(e^H)$  for all  $e^H \in A^+(H_0 : \zeta)$  that is equivalent to (81), which implies the proposition. Note that we take the differential operator  $b \in U(\mathfrak{g})$  appearing in that theorem to be trivial, in which case the operator  $\gamma_0(b)$  defined in (5.9.2) of [GV88] is also trivial.  $\square$

The second ingredient is the following estimate for the main term  $\theta_L$  in the asymptotic expansion (81).

**Proposition 9.4.** *For a standard Levi subgroup  $L$  write*

$$(82) \quad \Theta_L(H, \lambda) = \sum_{w \in W_L} \prod_{\alpha \in \Phi_{\text{red}, L}^+} f_\alpha(H, w\lambda)$$

*and write  $\Theta_G = \Theta$  to accord with (4). Assume Theorem 2.3 holds for  $L$  with corresponding constants  $a, \kappa$ , so that*

$$\varphi_\lambda^L(e^H) \ll (1 + \|\lambda\|)^a \Theta_L(H, \lambda) \max_{w \in W_L} e^{-(\rho_L + w\Im\lambda)(H)}$$

*for  $\lambda \in \mathfrak{a}^*(\kappa)$  and  $H \in \bar{\mathfrak{a}}_{L,+}$ , where  $\bar{\mathfrak{a}}_{L,+}$  is defined in (13). Then there are  $a', \kappa' > 0$  such that for all  $H \in \bar{\mathfrak{a}}_+ \cap \mathcal{C}_L$  and all  $\lambda \in \mathfrak{a}^*(\kappa')$ , we have*

$$\theta_L(H, \lambda) \ll (1 + \|\lambda\|)^{a'} \Theta(H, \lambda) \max_{w \in W} e^{-(\rho_L + w\Im\lambda)(H)}.$$

We shall prove Proposition 9.4 in the next paragraph. Let us show how to deduce Theorem 2.3 for  $H \in \bar{\mathfrak{a}}_+ \cap \mathcal{C}_L$  from these results. First, we apply the asymptotic expansion of  $\varphi_\lambda$  along  $L$  given in Proposition 9.3. The error term there is dominated by the majorant of Theorem 2.3, and Proposition 9.4 shows that the same is true of the main term, which completes the proof.

**9.6. Proof of Proposition 9.4.** Recall the definition of  $\mathfrak{a}_{\text{reg}}^*(\kappa)$  and  $\mathfrak{a}_{\text{reg}}^*(\kappa, C)$  from (78) and (79). We shall show that the case of  $\lambda \in \mathfrak{a}_{\text{reg}}^*(\kappa, c^{-1}\|H\|^{-1})$  can be treated in a straightforward way, using the hypothesis of the proposition. We shall then deduce the bound for all  $\lambda \in \mathfrak{a}^*(\kappa')$  for an appropriate choice of  $\kappa'$  from the case of  $\lambda \in \mathfrak{a}_{\text{reg}}^*(\kappa, c^{-1}\|H\|^{-1})$  using the maximum modulus principle, aided by Lemma 9.2.

For any  $\lambda \in \mathfrak{a}^*(\kappa)$ , we have the bound

$$(83) \quad c^L(\lambda) \ll \prod_{\alpha \in \Phi_{\text{red}}^+ \setminus \Phi_{\text{red},L}^+} (1 + |\langle \lambda, \alpha \rangle|^{-1}),$$

since for  $s$  small we have  $c_\alpha(s) \ll 1 + |s|^{-1}$  because of the simple pole at  $s = 0$ , and for  $s$  large with imaginary part bounded by  $\kappa$ , we have  $c_\alpha(s) \ll_\kappa |s|^{-m_\alpha/2 - m_{2\alpha}/2} \ll 1 + |s|^{-1}$ . Using (83) and the assumption of the proposition, we get

$$\theta_L(H, \lambda) \ll (1 + \|\lambda\|)^a \Psi_L(H, \lambda) \max_{w \in W} e^{-(\rho_L + w \Im \lambda)(H)}$$

for any  $\lambda \in \mathfrak{a}(\kappa)^*$ , where

$$\Psi_L(H, \lambda) = \sum_{w \in W_L \setminus W} \Theta_L(H, w\lambda) \prod_{\alpha \in \Phi_{\text{red}}^+ \setminus \Phi_{\text{red},L}^+} (1 + |\langle w\lambda, \alpha \rangle|^{-1}).$$

We claim that, when  $\lambda \in \mathfrak{a}_{\text{reg}}^*(\kappa, c^{-1}\|H\|^{-1})$ , we have  $\Psi_L(H, \lambda) \ll \Theta(H, \lambda)$ . To see this, first note that, by the  $W$ -invariance of  $\langle \cdot, \cdot \rangle$  and the  $W_L$ -invariance of  $\Phi_{\text{red}}^+ \setminus \Phi_{\text{red},L}^+$  (see Proposition 3.1), we have

$$\prod_{\alpha \in \Phi_{\text{red}}^+ \setminus \Phi_{\text{red},L}^+} (1 + |\langle w\lambda, \alpha \rangle|^{-1}) = \prod_{\alpha \in \Phi_{\text{red}}^+ \setminus \Phi_{\text{red},L}^+} (1 + |\langle sw\lambda, s\alpha \rangle|^{-1}) = \prod_{\alpha \in \Phi_{\text{red}}^+ \setminus \Phi_{\text{red},L}^+} (1 + |\langle sw\lambda, \alpha \rangle|^{-1}).$$

for any  $s \in W_L$ . Thus, recalling the definition of  $\Theta_L(H, w\lambda)$  in (82), we find

$$\begin{aligned} \Psi_L(H, \lambda) &= \sum_{w \in W_L \setminus W} \sum_{s \in W_L} \prod_{\alpha \in \Phi_{\text{red},L}^+} f_\alpha(H, sw\lambda) \prod_{\alpha \in \Phi_{\text{red}}^+ \setminus \Phi_{\text{red},L}^+} (1 + |\langle w\lambda, \alpha \rangle|^{-1}) \\ &= \sum_{w \in W_L \setminus W} \sum_{s \in W_L} \prod_{\alpha \in \Phi_{\text{red},L}^+} f_\alpha(H, sw\lambda) \prod_{\alpha \in \Phi_{\text{red}}^+ \setminus \Phi_{\text{red},L}^+} (1 + |\langle sw\lambda, \alpha \rangle|^{-1}) \end{aligned}$$

Note that for  $H \in \mathcal{C}_L$  and  $\lambda \in \mathfrak{a}_{\text{reg}}^*(\kappa, c^{-1}\|H\|^{-1})$ , we have  $|\langle \alpha, \lambda \rangle|^{-1} \leq c\|H\| < \alpha(H)$  for all  $\alpha \in \Phi_{\text{red}}^+ \setminus \Phi_{\text{red},L}^+$ , so that  $(1 + |\langle sw\lambda, \alpha \rangle|^{-1}) \ll f_\alpha(H, sw\lambda)$ . We put together the above estimates to obtain

$$\Psi_L(H, \lambda) \ll \sum_{w \in W_L \setminus W} \sum_{s \in W_L} \prod_{\alpha \in \Phi_{\text{red}}^+} f_\alpha(H, sw\lambda) = \sum_{s \in W} \prod_{\alpha \in \Phi_{\text{red}}^+} f_\alpha(H, s\lambda) = \Theta(H, \lambda).$$

The above bounds then establish the case of  $\lambda \in \mathfrak{a}_{\text{reg}}^*(\kappa; c^{-1}\|H\|^{-1})$ .

We now let  $\kappa' \in (0, \kappa)$  and take  $\lambda \in \mathfrak{a}^*(\kappa')$ . We apply Lemma 9.2 to  $\lambda$  and  $s = c^{-1}\|H\|^{-1}$ , and let  $C = C(H, \lambda) \in [1, \sigma]$  be the constant produced. Note that to apply Lemma 9.2, we may have had to make  $R$  bigger which we are free to do. Applying the maximum modulus principle, together with the already established bound in the case of  $\lambda \in \mathfrak{a}_{\text{reg}}^*(\kappa, c^{-1}\|H\|^{-1})$ , gives

$$\begin{aligned} \theta_L(H, \lambda) &\leq \max_{|z|=Cc^{-1}\|H\|^{-1}} \theta_L(H, \lambda + z\rho) \\ &\ll \max_{|z|=Cc^{-1}\|H\|^{-1}} (1 + \|\lambda + z\rho\|)^a \Theta(H, \lambda + z\rho) \max_{w \in W} e^{-(\rho_L + w\Im(\lambda + z\rho))(H)} \\ &\ll (1 + \|\lambda\|)^a \max_{w \in W} e^{-(\rho_L + w\Im(\lambda + z\rho))(H)} \max_{|z|=Cc^{-1}\|H\|^{-1}} \Theta(H, \lambda + z\rho). \end{aligned}$$

Note that in the above chain of inequalities it was crucial that, even though  $C(\lambda)$  varies with  $\lambda$ ,  $C(\lambda)$  always lies in a bounded range.

It therefore remains to show that

$$\max_{|z|=Cc^{-1}\|H\|^{-1}} \Theta(H, \lambda + z\rho) \ll \Theta(H, \lambda),$$

which in turn follows from

$$(84) \quad \min(|\alpha(H)| + 1, |\langle \lambda, \beta \rangle|^{-1} + 1) \asymp \min(|\alpha(H)| + 1, |\langle \lambda + z\rho, \beta \rangle|^{-1} + 1)$$

for  $|z| = Cc^{-1}\|H\|^{-1}$  and  $\alpha, \beta \in \Phi_{\text{red}}$ . In the case when  $|\langle \lambda, \beta \rangle| < c^{-1}\|H\|^{-1}$ , we have  $|\langle \lambda + z\rho, \beta \rangle| \ll \|H\|^{-1}$ , so that

$$|\langle \lambda, \beta \rangle|^{-1}, |\langle \lambda + z\rho, \beta \rangle|^{-1} \gg \|H\| \gg |\alpha(H)|.$$

This implies that both minima are  $\asymp |\alpha(H)| + 1$ , as required. When  $|\langle \lambda, \beta \rangle| \geq c^{-1}\|H\|^{-1}$ , we have  $|\langle \lambda, \beta \rangle| \asymp |\langle \lambda + z\rho, \beta \rangle|$ , which again implies (84).

This completes the proof of Proposition 9.4 and hence of Theorem 2.3.

**9.7. Proof of Theorem 2.4.** We now deduce Theorem 2.4 from Theorem 2.3. We will do this by passing between the bounds (6) and (7) for  $\varphi_\lambda^E$  and  $\varphi_\lambda$ , and so for the convenience of the reader we recall that these bounds state that

$$(85) \quad \varphi_\lambda^E(H) \ll \sum_{w \in W} \prod_{\alpha \in \Phi^+} (1 + |\alpha(H)\langle w\lambda, \alpha \rangle|)^{-1},$$

and

$$(86) \quad \varphi_\lambda(e^H) \ll e^{-\rho(H)} \prod_{\alpha \in \Phi^+} (|\alpha(H)| + 1) \sum_{w \in W} \prod_{\alpha \in \Phi^+} (1 + |\alpha(H)\langle w\lambda, \alpha \rangle|)^{-1},$$

for  $H \in \bar{\mathfrak{a}}^+$  and  $\lambda \in \mathfrak{a}^*$ . To establish these, Lemma 9.5 below shows that Theorem 2.3 implies (86) when  $\|\lambda\| < 1$ , which in turn implies (85) when  $\|\lambda\| < 1$ . We then use the fact that  $\varphi_{r\lambda}^E(H/r) = \varphi_\lambda^E(H)$  for any  $r > 0$ , which follows from the bilinearity of the inner product in the phase function in (5), to deduce (85) for all  $\lambda$ . This implies (86) for all  $\lambda$ , and completes the proof.

**Lemma 9.5.** *Let  $R > 0$  be given. In the range when  $\lambda \in \mathfrak{a}^*$  satisfies  $\|\lambda\| < R$ , the bounds for  $\varphi_\lambda$  given by Theorems 2.3 and 2.4 are equivalent up to a constant factor depending on  $R$ .*

*Proof.* If we define

$$g_\alpha(H, w\lambda) = \frac{|\alpha(H)| + 1}{|\alpha(H)\langle w\lambda, \alpha \rangle| + 1}$$

for  $H \in \mathfrak{a}$ ,  $\lambda \in \mathfrak{a}_\mathbb{C}^*$ , and  $w \in W$ , then the bound of Theorem 2.4 reads

$$\varphi_\lambda(e^H) \ll e^{-\rho(H)} \sum_{w \in W} \prod_{\alpha \in \Phi^+} g_\alpha(H, w\lambda).$$

It therefore suffices to prove that  $f_\alpha(H, w\lambda) \asymp_R g_\alpha(H, w\lambda)$  for  $\|\lambda\| < R$ , as this will imply that

$$\Theta(H, \lambda) = \sum_{w \in W} \prod_{\alpha \in \Phi^+} f_\alpha(H, w\lambda) \asymp_R \sum_{w \in W} \prod_{\alpha \in \Phi^+} g_\alpha(H, w\lambda)$$

as required. To establish  $f_\alpha \asymp g_\alpha$ , we consider the cases  $|\alpha(H)\langle w\lambda, \alpha \rangle| < 1$  and  $|\alpha(H)\langle w\lambda, \alpha \rangle| \geq 1$  separately. When  $|\alpha(H)\langle w\lambda, \alpha \rangle| < 1$ , we have

$$f_\alpha(H, w\lambda) = |\alpha(H)| + 1 \asymp g_\alpha(H, w\lambda),$$

as required. When  $|\alpha(H)\langle w\lambda, \alpha \rangle| \geq 1$ , we have

$$f_\alpha(H, w\lambda) = |\langle w\lambda, \alpha \rangle|^{-1} + 1$$

and

$$g_\alpha(H, w\lambda) \asymp \frac{|\alpha(H)| + 1}{|\alpha(H)\langle w\lambda, \alpha \rangle|} = |\langle w\lambda, \alpha \rangle|^{-1} + |\alpha(H)\langle w\lambda, \alpha \rangle|^{-1}.$$

The result follows in this case because  $|\langle w\lambda, \alpha \rangle|^{-1} \gg_R 1$  and  $|\alpha(H)\langle w\lambda, \alpha \rangle|^{-1} \leq 1$ .  $\square$

## REFERENCES

- [ABB+17] Miklos Abert, Nicolas Bergeron, Ian Biringer, Tsachik Gelander, Nikolay Nikolov, Jean Raimbault, and Iddo Samet. “On the growth of  $L^2$ -invariants for sequences of lattices in Lie groups”. In: *Ann. of Math.* (2) 185.3 (2017), pp. 711–790.
- [ABM22] Miklos Abert, Nicolas Bergeron, and Etienne Le Masson. *Eigenfunctions and Random Waves in the Benjamini-Schramm limit*. 2022. arXiv: [1810 . 05601 \[math.SP\]](https://arxiv.org/abs/1810.05601).
- [ALM15] Nalini Anantharaman and Etienne Le Masson. “Quantum ergodicity on large regular graphs”. In: *Duke Math. J.* 164.4 (2015), pp. 723–765.
- [Ank87] Jean-Philippe Anker. “La forme exacte de l'estimation fondamentale de Harish-Chandra”. In: *C. R. Acad. Sci. Paris Sér. I Math.* 305.9 (1987), pp. 371–374.
- [AS19] Nalini Anantharaman and Mostafa Sabri. “Quantum ergodicity on graphs: from spectral to spatial delocalization”. In: *Ann. of Math.* (2) 189.3 (2019), pp. 753–835.
- [Bar93] Alexander I. Barvinok. “Computing the volume, counting integral points, and exponential sums”. In: *Discrete Comput. Geom.* 10.2 (1993), pp. 123–141.
- [BC86] Daniel Barlet and Jean-Louis Clerc. “Le comportement à l'infini des fonctions de Bessel généralisées. I”. In: *Adv. in Math.* 61.2 (1986), pp. 165–183.
- [BLML16] Shimon Brooks, Etienne Le Masson, and Elon Lindenstrauss. “Quantum ergodicity and averaging operators on the sphere”. In: *Int. Math. Res. Not. IMRN* 19 (2016), pp. 6034–6064.

[BM23] Farrell Brumley and Jasmin Matz. “Quantum ergodicity for compact quotients of  $\mathrm{SL}_d(\mathbb{R})/\mathrm{SO}(d)$  in the Benjamini-Schramm limit”. In: *J. Inst. Math. Jussieu* 22.5 (2023), pp. 2075–2115.

[Bou02] Nicolas Bourbaki. *Lie groups and Lie algebras. Chapters 4–6*. Elements of Mathematics (Berlin). Translated from the 1968 French original by Andrew Pressley. Springer-Verlag, Berlin, 2002, pp. xii+300.

[Bri88] Michel Brion. “Points entiers dans les polyèdres convexes”. In: *Ann. Sci. École Norm. Sup. (4)* 21.4 (1988), pp. 653–663.

[Cle87] Jean-Louis Clerc. “Le comportement à l’infini des fonctions de Bessel généralisées. II”. In: *Adv. in Math.* 66.1 (1987), pp. 31–61.

[CN01] M. Cowling and A. Nevo. “Uniform estimates for spherical functions on complex semisimple Lie groups”. In: *Geom. Funct. Anal.* 11.5 (2001), pp. 900–932.

[Dei18] Anton Deitmar. “Benjamini-Schramm and spectral convergence”. In: *Enseign. Math.* 64.3-4 (2018), pp. 371–394.

[DKV79] J. J. Duistermaat, J. A. C. Kolk, and V. S. Varadarajan. “Spectra of compact locally symmetric manifolds of negative curvature”. In: *Invent. Math.* 52.1 (1979), pp. 27–93.

[GN15] Alexander Gorodnik and Amos Nevo. “Quantitative ergodic theorems and their number-theoretic applications”. In: *Bull. Amer. Math. Soc. (N.S.)* 52.1 (2015), pp. 65–113.

[GV88] Ramesh Gangolli and V. S. Varadarajan. *Harmonic analysis of spherical functions on real reductive groups*. Vol. 101. Ergebnisse der Mathematik und ihrer Grenzgebiete [Results in Mathematics and Related Areas]. Springer-Verlag, Berlin, 1988, pp. xiv+365.

[Hel00] Sigurdur Helgason. *Groups and geometric analysis*. Vol. 83. Mathematical Surveys and Monographs. Integral geometry, invariant differential operators, and spherical functions, Corrected reprint of the 1984 original. American Mathematical Society, Providence, RI, 2000, pp. xxii+667.

[Hel01] Sigurdur Helgason. *Differential geometry, Lie groups, and symmetric spaces*. Vol. 34. Graduate Studies in Mathematics. Corrected reprint of the 1978 original. American Mathematical Society, Providence, RI, 2001, pp. xxvi+641.

[Hum75] James E. Humphreys. *Linear algebraic groups*. Vol. No. 21. Graduate Texts in Mathematics. Springer-Verlag, New York-Heidelberg, 1975, pp. xiv+247.

[Ion00] Alexandru D. Ionescu. “An endpoint estimate for the Kunze-Stein phenomenon and related maximal operators”. In: *Ann. of Math. (2)* 152.1 (2000), pp. 259–275.

[KLM08] Michael Kapovich, Bernhard Leeb, and John J. Millson. “The generalized triangle inequalities in symmetric spaces and buildings with applications to algebra”. In: *Mem. Amer. Math. Soc.* 192.896 (2008), pp. viii+83.

[Kos73] Bertram Kostant. “On convexity, the Weyl group and the Iwasawa decomposition”. In: *Ann. Sci. École Norm. Sup. (4)* 6 (1973), pp. 413–455.

[KS09] Dubi Kelmer and Peter Sarnak. “Strong spectral gaps for compact quotients of products of  $\mathrm{PSL}(2, \mathbb{R})$ ”. In: *J. Eur. Math. Soc. (JEMS)* 11.2 (2009), pp. 283–313.

[KS60] R. A. Kunze and E. M. Stein. “Uniformly bounded representations and harmonic analysis of the  $2 \times 2$  real unimodular group”. In: *Amer. J. Math.* 82 (1960), pp. 1–62.

[LMS17] Etienne Le Masson and Tuomas Sahlsten. “Quantum ergodicity and Benjamini-Schramm convergence of hyperbolic surfaces”. In: *Duke Math. J.* 166.18 (2017), pp. 3425–3460.

[Nev98] Amos Nevo. “Spectral transfer and pointwise ergodic theorems for semi-simple Kazhdan groups”. In: *Math. Res. Lett.* 5.3 (1998), pp. 305–325.

[NPP14] E. K. Narayanan, A. Pasquale, and S. Pusti. “Asymptotics of Harish-Chandra expansions, bounded hypergeometric functions associated with root systems, and applications”. In: *Adv. Math.* 252 (2014), pp. 227–259.

[Pet23a] Carsten Peterson. *Quantum Ergodicity on Bruhat-Tits Buildings*. Thesis (Ph.D.)—University of Michigan. ProQuest LLC, Ann Arbor, MI, 2023, p. 180.

[Pet23b] Carsten Peterson. *Quantum ergodicity on the Bruhat-Tits building for  $PGL(3, F)$  in the Benjamini-Schramm limit*. 2023. arXiv: [2304.08641 \[math.RT\]](https://arxiv.org/abs/2304.08641).

[Pet24] Carsten Peterson. *A degenerate version of Brion’s formula*. 2024. arXiv: [2409.09544 \[math.CO\]](https://arxiv.org/abs/2409.09544).

[Sni74] A. I. Snirelman. “Ergodic properties of eigenfunctions”. In: *Uspehi Mat. Nauk* 29.6(180) (1974), pp. 181–182.

[Ste+23] W. A. Stein et al. *Sage Mathematics Software (Version 10.1)*. <http://www.sagemath.org>. The Sage Development Team. 2023.

[SV07] Lior Silberman and Akshay Venkatesh. “On quantum unique ergodicity for locally symmetric spaces”. In: *Geom. Funct. Anal.* 17.3 (2007), pp. 960–998.

[Ver85] Y. Colin de Verdière. “Ergodicité et fonctions propres du laplacien”. In: *Comm. Math. Phys.* 102.3 (1985), pp. 497–502.

[Zel87] Steven Zelditch. “Uniform distribution of eigenfunctions on compact hyperbolic surfaces”. In: *Duke Math. J.* 55.4 (1987), pp. 919–941.

SORBONNE UNIVERSITÉ

*Email address:* [brumley@imj-prg.fr](mailto:brumley@imj-prg.fr)

UNIVERSITY OF WISCONSIN

*Email address:* [marshall@math.wisc.edu](mailto:marshall@math.wisc.edu)

UNIVERSITY OF COPENHAGEN

*Email address:* [matz@math.ku.dk](mailto:matz@math.ku.dk)

SORBONNE UNIVERSITÉ

*Email address:* [peterson@imj-prg.fr](mailto:peterson@imj-prg.fr)