THE DENSITY HYPOTHESIS FOR HORIZONTAL FAMILIES OF LATTICES

MIKOŁAJ FRĄCZYK, GERGELY HARCOS, PÉTER MAGA, AND DJORDJE MILIĆEVIĆ

To Miklós Abért for bringing us together.

ABSTRACT. We prove the density hypothesis for wide families of arithmetic orbifolds arising from all division quaternion algebras over all number fields of bounded degree. Our power-saving bounds on the multiplicities of non-tempered representations are uniform in the volume and spectral aspects.

1. INTRODUCTION

1.1. **Exceptional spectrum.** Selberg's celebrated Eigenvalue Conjecture states that all nonzero Laplacian eigenvalues on congruence quotients of the upper half-plane are at least $1/4$. This is a strong form of the "spectral gap" property of the same eigenvalues being uniformly bounded away from 0. In representation-theoretic language, Selberg's conjecture states that the archimedean constituents of nontrivial automorphic representations occurring in the spherical discrete L^2-spectrum for congruence subgroups of $\mathrm{SL}_2(\mathbb{Z})$ are tempered, that is, no complementary series representations occur. It is thus the archimedean counterpart of the Ramanujan–Petersson conjecture stating that the suitably normalized Hecke eigenvalues of cusp forms are bounded on primes, and is expected to hold for more general Lie groups and their congruence quotients. We remark that the Ramanujan–Petersson conjecture needs to be adjusted for general groups, as explained in [52, §2].

Selberg's conjecture (as well as the Ramanujan conjecture for Maaß forms) remains far from being resolved. For individual forms the best known results are lower bounds on the spectral gap such as [36]. For analytic applications in a *family* of automorphic forms, in the absence of Selberg's conjecture, the non-tempered spectrum can often be satisfactorily handled if the exceptions in the family are known to be *sparse and not too bad*. The situation is reminiscent of prime number theory, where a classical density theorem for Dirichlet L-functions $L(s, \chi)$ estimates the total number of zeros $\beta + i\gamma$ with $\beta \geqslant \sigma \geqslant 1/2$, $|\gamma| \leqslant T$, and χ modulo q by $O_\varepsilon((qT)^{c(1-\sigma)+\varepsilon})$. Such a bound, conjectured with $c = 2$ [32, §10.1] and known with $c = 12/5$ [32, §10.4], serves (in the light of explicit formulae) as a good substitute for the Riemann Hypothesis in applications such as counting primes in short intervals or arithmetic progressions.

It is imperative to understand which families admit an analogous density estimate for the exceptional spectrum, such as the "correct" power-saving bound (cf. (1.4) and (1.13)) for the total multiplicity of Laplacian eigenvalues $1/4 + (i\nu)^2$ with $\nu \geqslant \sigma \geqslant 0$ on congruence surfaces of increasing level or surfaces arising from Eichler orders in varying division quaternion algebras. Our main result, Theorem 1 below, covers these among many other examples and proves for the first time the density hypothesis for a natural broad "horizontal" family of not necessarily commensurable arithmetic orbifolds, including also the spectral aspect.

1.2. **Density theorems, uniformity, and implications.** We now state the density hypothesis for a family of lattices in a semisimple Lie group and discuss its applications and connections to the limit multiplicity problem. We refer to [3] for an overview of the density conjecture with a somewhat different focus.

1.2.1. *Density conjecture and density hypothesis.* Let G be a semisimple real Lie group without compact factors, and let k be a number field with ring of integers $\mathfrak{o}$. Let $\mathbf{G}$ be a semisimple linear algebraic group over k such that $\mathbf{G}(k \otimes_{\mathbb{Q}} \mathbb{R})$ is isomorphic to G times a compact Lie group, and let ι denote the projection onto G. We identify $\mathbf{G}(k)$ with its image under a faithful linear representation $\mathbf{G}(k) \hookrightarrow \mathrm{GL}_n(k)$ and also with its image under the natural embedding $\mathbf{G}(k) \hookrightarrow \mathbf{G}(k \otimes_{\mathbb{Q}} \mathbb{R})$. We define the *principal congruence subgroups* as

$$\Gamma(\mathfrak{n}) := \iota\big(\{g \in \mathbf{G}(k) \cap \mathrm{GL}_n(\mathfrak{o}) : g - \mathrm{id}_n \in \mathfrak{n} M_n(\mathfrak{o})\}\big) \leqslant G,$$

where $\mathfrak{n} \subset \mathfrak{o}$ is a nonzero ideal. We refer to k as the *trace field* of $\Gamma(\mathfrak{o})$.

Let $\pi \in \widehat{G}$ be an irreducible unitary representation of G. We assign to it the real number

$$(1.1) \qquad p(\pi) := \inf\big\{2 \leqslant p \leqslant \infty : \pi \text{ has a nonzero matrix coefficient in } L^p(G)\big\}.$$

We recall that, for $p > p(\pi)$, all smooth matrix coefficients lie in $L^p(G)$. (Indeed, there exist nonzero $v, w \in V_\pi$ such that $c_{v,w}(g) := \langle \pi(g)v, w \rangle$ lies in $L^q(G)$ for some $p(\pi) \leqslant q < p$. As $c_{v,w}$ is bounded, it lies in $L^p(G)$ as well. By Jensen's inequality, the property that $c_{v,w} \in L^p(G)$ is invariant under the action of $L^1(G)$ on v and w. Hence we can assume that $v, w \in V_\pi^\infty$ are nonzero smooth vectors, and then [55, Th. 11.8.2] implies that $c_{v',w'} \in L^p(G)$ holds for all $v', w' \in V_\pi^\infty$.) A unitary representation π is tempered if and only if $p(\pi) = 2$. On the opposite side is the trivial representation $\mathbf{1}$ for which $p(\mathbf{1}) = \infty$. Using Harish-Chandra's expansion for matrix coefficients [37, Ch. VIII, §8] and the Howe–Moore vanishing theorem [26, Th. 5.1], it is easy to see that for G simple the trivial representation is uniquely characterized by this property.

For every lattice $\Gamma \leqslant G$, let $\mathrm{m}(\pi, \Gamma)$ be the multiplicity of a given $\pi \in \widehat{G}$ in the right regular representation of G on $L^2(\Gamma \backslash G)$. Sarnak and Xue conjectured [53, Conj. 1] that

$$(1.2) \qquad \mathrm{m}(\pi, \Gamma(\mathfrak{n})) \ll_\varepsilon \mathrm{vol}(\Gamma(\mathfrak{n}) \backslash G)^{2/p(\pi)+\varepsilon}.$$

This bound interpolates, for non-tempered representations, between the bound $\mathrm{m}(\pi, \Gamma(\mathfrak{n})) \ll \mathrm{vol}(\Gamma(\mathfrak{n}) \backslash G)$, which (at least for compact quotients) follows fairly directly from the trace formula and is sharp when $\pi \in \widehat{G}$ belongs to discrete series [11], and the identity $\mathrm{m}(\mathbf{1}, \Gamma(\mathfrak{n})) = 1$. Sarnak and Xue [53] proved that (1.2) holds for cocompact principal congruence lattices in $G = \mathrm{SL}_2(\mathbb{R})$ and $\mathrm{SL}_2(\mathbb{C})$. As pointed out by Sarnak [51, Remarks 2.4], (1.2) combined with Matsushima's formula gives rather strong bounds for the Betti numbers of cocompact congruence lattices $\Gamma(\mathfrak{n})$ (see [7, Ch. VII, §3.2], [7, Ch. II, §3.1], [58, (1)]), while the analogue of (1.2) for general (noncongruence) lattices of G is false.

Inequality (1.2) admits a natural strengthening. Let $\mathcal{B} \subset \widehat{G}$ be such that the infinitesimal characters of $\pi \in \mathcal{B}$ form a bounded set (cf. [37, Ch. VIII, §6]), and let us write

$$(1.3) \qquad p(\mathcal{B}) := \inf\{p(\pi) : \pi \in \mathcal{B}\}.$$

Then the *density conjecture* (whose origins go back to [51, 53]) states that, for every $\mathcal{B} \subset \widehat{G}$ as above,

$$(1.4) \qquad \sum_{\pi \in \mathcal{B}} \mathrm{m}(\pi, \Gamma) \ll_{\mathcal{B},\varepsilon} \mathrm{vol}(\Gamma \backslash G)^{2/p(\mathcal{B})+\varepsilon}$$

holds uniformly for all $\Gamma = \Gamma(\mathfrak{n})$. It is known for $G = \mathrm{SL}_2(\mathbb{R})$ and $G = \mathrm{SL}_2(\mathbb{C})$ by [28], and also in further SL_2 families of congruence quotients of a fixed lattice by [8, 27, 29, 30, 50] and others. Many partial results are known for other groups (see [3] as well as [4, 9, 34, 42]), although the full conjecture (including the individual bound (1.2)) is still open if G is any simple group of higher rank.

More generally, following [23, 24], we say that a *family* of lattices $\Gamma \leqslant G$ satisfies the *density hypothesis* if (1.4) holds uniformly for lattices in that family.

Recently, Golubev and Kamber [23, 24] applied the Sarnak–Xue conjecture in various combinatorial contexts. Among their applications is the so-called *optimal lifting property* [24, Def. 1.4], which says (in our setting) that almost all elements of the quotient group $\Gamma(\mathfrak{n}) \backslash \Gamma(\mathfrak{o})$ can be lifted to elements of $\Gamma(\mathfrak{o})$ lying in a ball of volume roughly $\mathrm{vol}(\Gamma(\mathfrak{n}) \backslash G)$; clearly, no smaller ball suffices. By [24, Thm. 1.6], which applies to more general families of lattices with a spectral gap, the optimal lifting property for the family $\Gamma(\mathfrak{n})$ is implied by the following *spherical density hypothesis*, a somewhat stronger version of (1.4) that addresses the dependence of the implicit constant on $\mathcal{B}$. Let $\widehat{G}_{\mathrm{sph}} \subset \widehat{G}$ be the spherical unitary dual of G, and let $\Omega(\pi)$ be the eigenvalue of the Casimir operator acting on $\pi \in \widehat{G}$. Following [24, Def. 1.2] we say that a family of

lattices $\mathcal{F}$ satisfies the spherical density hypothesis if there exists $L > 0$ such that for every $\mathcal{B} \subset \widehat{G}_{\mathrm{sph}}$ with $T := \sup_{\pi \in \mathcal{B}} \Omega(\pi) < \infty$ and every lattice $\Gamma \in \mathcal{F}$ we have

$$(1.5) \qquad \sum_{\pi \in \mathcal{B}} \mathrm{m}(\pi, \Gamma) \ll_\varepsilon \mathrm{vol}(\Gamma \backslash G)^{2/p(\mathcal{B}) + \varepsilon} (1 + T)^L.$$

Our main result, Theorem 1 below, establishes this hypothesis for a wide family of lattices in $G = \mathrm{SL}_2(\mathbb{R})^a \times \mathrm{SL}_2(\mathbb{C})^b$.

1.2.2. *Limit multiplicity property.* Our extension of Sarnak–Xue's results was motivated by recent developments in the limit multiplicity property. Let $\Gamma \leqslant G$ be a lattice. The right regular representation $L^2(\Gamma \backslash G)$ decomposes into discrete part $L^2(\Gamma \backslash G)_{\mathrm{disc}}$ and continuous part $L^2(\Gamma \backslash G)_{\mathrm{cont}}$. We have

$$L^2(\Gamma \backslash G)_{\mathrm{disc}} = \widehat{\bigoplus_{\pi \in \widehat{G}}} \mathrm{m}(\pi, \Gamma) \pi.$$

Let $\mathcal{F}$ be a family of lattices in G intersecting the center of G in the same subgroup Θ. Assume that arbitrary large covolumes occur in $\mathcal{F}$, and denote by $\mu_{\widehat{G/\Theta}}$ the Plancherel measure on the unitary dual of G/Θ. We say that $\mathcal{F}$ has the *limit multiplicity property* if the measures

$$\mu_\Gamma := \frac{1}{\mathrm{vol}(\Gamma \backslash G)} \sum_{\pi \in \widehat{G/\Theta}} \mathrm{m}(\pi, \Gamma) \delta_\pi, \qquad \Gamma \in \mathcal{F},$$

tend to $\mu_{\widehat{G/\Theta}}$ as $\mathrm{vol}(\Gamma \backslash G) \to \infty$ in the following sense:

- $\mu_\Gamma(\mathcal{B}) \to \mu_{\widehat{G/\Theta}}(\mathcal{B})$ for every Jordan measurable subset $\mathcal{B} \subset \widehat{G/\Theta}$ of tempered representations;

- $\mu_\Gamma(\mathcal{B}) \to 0$ for every subset $\mathcal{B} \subset \widehat{G/\Theta}$ of non-tempered representations whose infinitesimal characters form a bounded set.

The limit multiplicity property was discovered by DeGeorge and Wallach [11, 12], and it implies a small-o version of (1.4). That the distribution of $\pi \in \widehat{G/\Theta}$ occurring in $L^2(\Gamma \backslash G)$ as $\mathrm{vol}(\Gamma \backslash G) \to \infty$ should be guided by the Plancherel measure on $\widehat{G/\Theta}$ may also be thought of as the lattice (in particular, level) counterpart of the same property for the leading term in Weyl's law.

Until quite recently, the limit multiplicity property was only studied for families of principal congruence subgroups of a fixed arithmetic lattice. The state-of-the-art in this setting is due to Finis, Lapid and Müller [18, 17]. The situation changed when Abért et al. proved [1, Th. 1.2] that if G is a simple group of higher rank, then the limit multiplicity property holds for any family of lattices $\mathcal{F}$ satisfying the following condition:

there exists an open set $U \subset G$ such that $g \Gamma g^{-1} \cap U = \{1\}$ for every $\Gamma \in \mathcal{F}$ and $g \in G$.

This is automatically satisfied if $\mathcal{F}$ consists of cocompact torsion-free arithmetic lattices whose trace field is of bounded degree over $\mathbb{Q}$ (see [22, §10] and [16, Th. 1]). Using other methods, the first named author proved in [19] that the family of all cocompact torsion-free congruence lattices of $\mathrm{SL}_2(\mathbb{R})$ and $\mathrm{SL}_2(\mathbb{C})$ has the limit multiplicity property.

1.3. **Main result.** Our main result, Theorem 1 below, establishes a strong form of the density hypothesis for natural broad "horizontal" families of arithmetic lattices, arising from all suitable division quaternion algebras over varying number fields. We describe these families of lattices in §1.3.1 and §1.3.2, the spherical unitary representations entering (1.4) in §1.3.3, and then state Theorem 1 in §1.3.4.

1.3.1. *Arithmetic lattices and orbifolds.* We begin with a standard construction of arithmetic lattices using quaternion algebras.

Let $a, b, c \in \mathbb{N}$ with $a + b \geqslant 1$. Let k be a number field of degree $d = a + 2b + c$ with b complex places. We enumerate the archimedean completions of k by $k_1, \ldots, k_{a+b+c}$, with the complex ones being $k_{a+1}, \ldots, k_{a+b}$. Let A be a quaternion algebra over k of signature (a, b, c), so that $A \otimes_{\mathbb{Q}} \mathbb{R} \simeq \mathrm{M}_2(\mathbb{R})^a \times \mathrm{M}_2(\mathbb{C})^b \times \mathbb{H}^c$, where $\mathbb{H} = \left(\frac{-1, -1}{\mathbb{R}} \right)$ stands for the Hamilton quaternion algebra. Let n (resp. tr) be the reduced norm (resp. reduced trace) on A. We will write $\mathrm{ram}(A)$ for the set of places of k where A ramifies.

Let $\mathbf{G} := \mathrm{SL}_1(A)$ viewed as an algebraic group over k, and let $G := \mathrm{SL}_2(\mathbb{R})^a \times \mathrm{SL}_2(\mathbb{C})^b$. For every k-algebra R we have

$$\mathbf{G}(R) = \{x \in A \otimes_k R : \mathrm{n}(x) = 1\}.$$

In particular, writing $\mathbb{A} = \mathbb{A}_\infty \times \mathbb{A}_f$ for the adele ring of k, we have

$$\mathbf{G}(\mathbb{A}_\infty) \simeq \mathbf{G}(k \otimes_\mathbb{Q} \mathbb{R}) \simeq \{x \in A \otimes_\mathbb{Q} \mathbb{R} : \mathrm{n}(x) = 1\} \simeq G \times \mathrm{SU}_2(\mathbb{C})^c,$$

and hence

(1.6) $$\mathbf{G}(\mathbb{A}) \simeq G \times \mathrm{SU}_2(\mathbb{C})^c \times \mathbf{G}(\mathbb{A}_f).$$

Write $\eta \colon \mathbf{G}(\mathbb{A}) \to G$ for the projection onto the first factor in (1.6). It is determined uniquely up to automorphisms of G. For a compact open subgroup $U \leqslant \mathbf{G}(\mathbb{A}_f)$, the group

$$\Gamma_U := \eta\big(\mathbf{G}(k) \cap (\mathbf{G}(\mathbb{A}_\infty) \times U)\big) \leqslant G$$

is a congruence lattice of G with invariant trace field $k\Gamma_U = k$ and invariant quaterion algebra $A\Gamma_U = A$ as defined in [41, Def. 3.3.6]. (Indeed, clearly, $k\Gamma_U \leqslant k$. One can see that a given $r \in k$ lies in $k\Gamma_U$ as follows. For a suitable positive integer n, there exist four elements in $\mathbf{G}(k) \cap (\mathbf{G}(\mathbb{A}_\infty) \times U)$ whose reduced traces are of the shape t, $t + n$, $t + nr$, $t + n + nr$. Hence t^2, $(t + n)^2$, $(t + nr)^2$, $(t + n + nr)^2$ lie in $k\Gamma_U$, while r lies in the $\mathbb{Q}$-span of these elements. Since $A\Gamma_U \subset A$ is a quaternion algebra over its center $k\Gamma_U = k$, it also follows that $A\Gamma_U = A$.) Hence the commensurability class of Γ_U determines the pair (k, A), and vice versa [41, Th. 3.3.4 & Cor. 3.3.5].

Let K be a maximal compact subgroup of G. Then, $G/K \simeq (\mathcal{H}^2)^a \times (\mathcal{H}^3)^b$, and its quotient $\Gamma_U \backslash G/K$ is an arithmetic orbifold, which is compact if and only if A is a division quaternion algebra (as will be the case in our Theorem 1). This description includes, for $G = \mathrm{SL}_2(\mathbb{R})$ and $G = \mathrm{SL}_2(\mathbb{C})$, all arithmetic hyperbolic 2- and 3-orbifolds.

1.3.2. *Families of congruence lattices.* For a nonzero ideal $\mathfrak{n} \subset \mathfrak{o}$ not divisible by any $\mathfrak{p} \in \mathrm{ram}(A)$, we introduce certain natural congruence lattices $\Gamma_\kappa(\mathfrak{n}) \leqslant G$ closely related to the classical congruence subgroups $\Gamma_0(n), \Gamma(n) \leqslant \mathrm{SL}_2(\mathbb{Z})$. Fix an isomorphism

(1.7) $$\mathbf{G}(\mathbb{A}_f) \simeq \prod_{\mathfrak{p} \in \mathrm{ram}(A)} \mathrm{SL}_1(D_\mathfrak{p}) \times \prod_{\mathfrak{p} \notin \mathrm{ram}(A)} \mathrm{SL}_2(k_\mathfrak{p}),$$

where $D_\mathfrak{p}$ is the unique non-split quaternion algebra over $k_\mathfrak{p}$. The group $\mathrm{SL}_1(D_\mathfrak{p})$ is anisotropic over $k_\mathfrak{p}$ (i.e., it has no split tori), so it is compact by [6, §9.4]; see also [5, §6.4]. Writing $\mathfrak{n}_\mathfrak{p} := \mathfrak{n}\mathfrak{o}_\mathfrak{p}$, consider the local compact open subgroups $K_0(\mathfrak{n}_\mathfrak{p}), K_1(\mathfrak{n}_\mathfrak{p}) \leqslant \mathrm{SL}_2(\mathfrak{o}_\mathfrak{p})$ defined as in (4.1)–(4.2) by

$$K_0(\mathfrak{n}_\mathfrak{p}) := \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathfrak{o}_\mathfrak{p}) : c \in \mathfrak{n}_\mathfrak{p} \right\}, \qquad K_1(\mathfrak{n}_\mathfrak{p}) := \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathfrak{o}_\mathfrak{p}) : a - d, b, c \in \mathfrak{n}_\mathfrak{p} \right\}$$

For any $\kappa \colon \{\mathfrak{p} \mid \mathfrak{n}\} \to \{0, 1\}$, define the compact open subgroup $K_\kappa(\mathfrak{n}) \subset \mathbf{G}(\mathbb{A}_f)$ as

(1.8) $$K_\kappa(\mathfrak{n}) := \prod_{\mathfrak{p} \in \mathrm{ram}(A)} \mathrm{SL}_1(D_\mathfrak{p}) \times \prod_{\mathfrak{p} \mid \mathfrak{n}} K_{\kappa(\mathfrak{p})}(\mathfrak{n}_\mathfrak{p}) \times \prod_{\substack{\mathfrak{p} \notin \mathrm{ram}(A) \\ \mathfrak{p} \nmid \mathfrak{n}}} \mathrm{SL}_2(\mathfrak{o}_\mathfrak{p}),$$

so that in particular

(1.9) $$K_\emptyset(\mathfrak{o}) = \prod_{\mathfrak{p} \in \mathrm{ram}(A)} \mathrm{SL}_1(D_\mathfrak{p}) \times \prod_{\mathfrak{p} \notin \mathrm{ram}(A)} \mathrm{SL}_2(\mathfrak{o}_\mathfrak{p}).$$

Define the corresponding congruence lattice $\Gamma_\kappa(\mathfrak{n}) \leqslant G$ as

(1.10) $$\Gamma_\kappa(\mathfrak{n}) := \Gamma_{K_\kappa(\mathfrak{n})}.$$

We remark that $\Gamma_\kappa(\mathfrak{n})$ depends on the choice of the identification (1.7), but our estimates will not depend on this choice. We are now ready to declare the family of lattices which Theorem 1 is concerned with.

Definition 1. For $a, b, c \in \mathbb{N}$, let $\mathcal{F}_{a,b,c}$ be the family of all congruence lattices $\Gamma_\kappa(\mathfrak{n}) \leqslant \mathrm{SL}_2(\mathbb{R})^a \times \mathrm{SL}_2(\mathbb{C})^b$ as in (1.10), where k is a number field of degree $a + 2b + c$, A is a division quaternion algebra over k of signature (a, b, c), $\mathfrak{n} \subset \mathfrak{o}_k$ is a nonzero ideal, and $\kappa \colon \{\mathfrak{p} \mid \mathfrak{n}\} \to \{0, 1\}$.

Thus $\mathcal{F}_{a,b,c}$ is a family of cocompact congruence lattices in G, which consists of infinitely many different commensurability classes in bijection with different pairs (k, A), by the discussion above.

1.3.3. *Representations.* Let $S_\infty^G = \{1, \ldots, a+b\}$. The spherical unitary dual $\widehat{G}_{\mathrm{sph}}$ can be parametrized as $\pi_{\boldsymbol{s}}$ with tuples $\boldsymbol{s} = (s_j)$ satisfying $s_j \in (0, 1/2] \cup i[0, \infty)$ for all $j \in S_\infty^G$ as follows. For $s \in \mathbb{C}$ and $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$, let $\pi_s^{\mathbb{K}}$ be the spherical principal series representation of $\mathrm{SL}_2(\mathbb{K})$ of normalized Casimir eigenvalue $1/4 - s^2$. In particular, $\pi_{-s}^{\mathbb{K}} = \pi_s^{\mathbb{K}}$. The point $s = 1/2$ corresponds to the trivial representation, the interval $(0, 1/2)$ corresponds to the non-tempered nontrivial spherical unitary spectrum, and the half-line $i[0, \infty)$ corresponds to the tempered spherical unitary spectrum. We give more background on spherical representations in §3.3. For $\boldsymbol{s} = (s_j)_{j=1}^{a+b}$, we define $\pi_{\boldsymbol{s}} := \bigotimes_{j=1}^{a+b} \pi_{s_j}^{k_j}$. It lies in $\widehat{G}$ if and only if each s_j lies in $[-1/2, 1/2] \cup i\mathbb{R}$, and in $\widehat{G}_{\mathrm{temp}}$ if and only if each s_j lies in $i\mathbb{R}$. Using [37, Th. 8.48], we may verify that (cf. (1.1))

$$(1.11) \qquad p(\pi_{\boldsymbol{s}}) = \max_{j \in S_\infty^G} p(\pi_{s_j}^{k_j}) = 2/\min_{j \in S_\infty^G}(1 - 2|\mathrm{Re}(s_j)|), \qquad \pi_{\boldsymbol{s}} \in \widehat{G}.$$

For $S \subset S_\infty^G$, $\boldsymbol{\sigma} = (\sigma_j) \in [0, 1/2]^S$, and $\boldsymbol{T} = (T_j) \in \mathbb{R}^{S_\infty^G \setminus S}$, we introduce the following bounded subset of $\widehat{G}_{\mathrm{sph}}$:

$$(1.12) \qquad \mathcal{B}(\boldsymbol{\sigma}, \boldsymbol{T}) := \left\{ \pi_{\boldsymbol{s}} \; : \; \boldsymbol{s} \in \prod_{j \in S}[\sigma_j, 1/2] \times \prod_{j \in S_\infty^G \setminus S} i[T_j - 1, T_j + 1] \right\}.$$

It is clear from (1.3) and (1.11) that

$$(1.13) \qquad p(\mathcal{B}(\boldsymbol{\sigma}, \boldsymbol{T})) = p(\boldsymbol{\sigma}) := \begin{cases} 2/\min_{j \in S}(1 - 2\sigma_j), & \text{if } S \neq \emptyset, \\ 2, & \text{if } S = \emptyset. \end{cases}$$

1.3.4. *Statement of the main result.* Let $\rho_j := 1$ for $j \in \{1, \ldots, a\}$ and $\rho_j := 2$ for $j \in \{a+1, \ldots, a+b\}$. In the spirit of the analytic conductor of Iwaniec–Sarnak [33], we define for an arbitrary lattice $\Gamma \in \mathcal{F}_{a,b,c}$ and tuple $\boldsymbol{T} \in \mathbb{R}^{S_\infty^G \setminus S}$ the quantity

$$(1.14) \qquad \mathcal{C}(\Gamma, \boldsymbol{T}) := \mathrm{vol}(\Gamma \backslash G) \prod_{j \in S_\infty^G \setminus S} (1 + |T_j|)^{\rho_j}.$$

We note that $\mathrm{vol}(\Gamma \backslash G)$ always exceeds e^{-7} by Borel's volume formula and Odlyzko's discriminant bound (see Proposition 4). Recalling Definition 1 and (1.12)–(1.14), our main result reads as follows.

Theorem 1. *For every $a, b, c \in \mathbb{N}$, the family $\mathcal{F}_{a,b,c}$ of congruence lattices in $G = \mathrm{SL}_2(\mathbb{R})^a \times \mathrm{SL}_2(\mathbb{C})^b$ satisfies the spherical density hypothesis. More precisely, for every $\Gamma \in \mathcal{F}_{a,b,c}$, $S \subset S_\infty^G$, $\boldsymbol{\sigma} \in [0, 1/2]^S$, and $\boldsymbol{T} \in \mathbb{R}^{S_\infty^G \setminus S}$, we have for any $\varepsilon > 0$*

$$(1.15) \qquad \sum_{\pi \in \mathcal{B}(\boldsymbol{\sigma}, \boldsymbol{T})} \mathrm{m}(\pi, \Gamma) \ll_{\varepsilon, a, b, c} \mathcal{C}(\Gamma, \boldsymbol{T})^{2/p(\boldsymbol{\sigma}) + \varepsilon}.$$

Remark 1. Our method is capable of handling larger families. Namely, assume that in (1.12) we replace each interval $[T_j - 1, T_j + 1]$ by $[T_j - W_j, T_j + W_j]$, where $W_j \in [1, T_j]$. Assume also that we multiply the right-hand side of (1.14) by the product of the W_j's. With these modifications, the bound (1.15) remains valid. To prove this more general result, one simply needs to divide $s \pm it$ in (3.27) by a new parameter $w \in [1, t]$, and update Proposition 2 accordingly. From this perspective, the quantity $\mathcal{C}(\Gamma, \boldsymbol{T})$, which in §1.4 we interpret as essentially the analytic conductor of representations in $\mathcal{B}(\boldsymbol{\sigma}, \boldsymbol{T})$ in the context of Theorem 1, can alternatively be understood as the expected order of magnitude of the left-hand side of (1.15) when $p(\boldsymbol{\sigma}) = 2$, i.e., when $\sigma_j = 0$ for all $j \in S$. We stick to the case $W_j = 1$ and $w = 1$ in Theorem 1 for notational simplicity.

Already for $S = S_\infty^G$ (or for fixed $\boldsymbol{T} \in \mathbb{R}^{S_\infty^G \setminus S}$), Theorem 1 extends for the first time the results of Sarnak–Xue [53] to families of non-commensurable lattices of G. Theorem 1 allows for groups G of arbitrary rank $a + b$ and a wider variety of congruence subgroups $\Gamma_\kappa(\mathfrak{n})$, even when considering subgroups of a fixed lattice. Moreover, it holds uniformly over all lattices in $\mathcal{F}_{a,b,c}$ and all possible pairs (k, A), with no dependence on a particular fixed ambient lattice, and addresses for the first time the dependence on the tempered components of π. In the fully degenerate case $S = \emptyset$, Theorem 1 recovers up to $\mathrm{vol}(\Gamma \backslash G)^\varepsilon$ the bound resulting

from Weyl's law for the group G (cf. (6.30) and [9, Prop. 7.3]), for the first time uniformly across all lattices in $\mathcal{F}_{a,b,c}$. Finally, (1.15) is in fact significantly stronger than (1.5) specialized to $\mathcal{F}_{a,b,c}$, in that it exhibits a natural dependence of the archimedean parameters (cf. discussion on $\mathcal{C}(\Gamma, \boldsymbol{T})$ in §1.4).

The dependence of the implicit constant in Theorem 1 on the signature (a, b, c), or equivalently on the degree $[k : \mathbb{Q}]$, seems difficult to remove. In particular, parts of our argument that appeal to the geometry of ideal lattices of k are very sensitive on the degree. This contrasts with the limit multiplicity property for all cocompact torsion-free congruence lattices of $\mathrm{SL}_2(\mathbb{R})$ and $\mathrm{SL}_2(\mathbb{C})$, obtained in [19] without any restriction on $[k : \mathbb{Q}]$.

1.4. **New features.** We shall be working with *strongly positive definite* functions $f \in C_c(G)$ (see §3.4), for which in particular $\pi(f)$ is a positive operator for every $\pi \in \widehat{G}$. Theorem 1 will be proved by the Arthur–Selberg trace formula, with a test function built from a strongly positive definite $f \in C_c(G)$, chosen so as to emphasize contributions of π occurring on the left-hand side of (1.15). We give a detailed overview of the method in §2, contending ourselves here with just pointing out several new features. The first is the *self-normalizing* nature of our estimates. Indeed, after the application of the trace formula with our specific test function, we may bound the left-hand side of (1.15) by (see §1.6 for notations, §2 and §6 for more details)

$$(1.16) \qquad e^{(2/p(\boldsymbol{\sigma})-1)R} \sum_{[\gamma] \subset \mathbf{G}(k)} \mathrm{vol}(\mathbf{G}_\gamma(k) \backslash \mathbf{G}_\gamma(\mathbb{A})) \cdot \mathbf{O}(\gamma, f) \mathbf{O}(\gamma, \mathbf{1}_{\mathrm{SU}_2(\mathbb{C})^c}) \mathbf{O}(\gamma, \mathbf{1}_{K_\kappa(\mathfrak{n})}).$$

In this expression, we estimate the volumes building on the work of Borel [5], Ono [46, 47], and Ullmo–Yafaev [54] (see Propositions 4–5); the orbital integrals using various integral transforms and counting in Bruhat–Tits trees (see Propositions 1–3); the number of contributing conjugacy classes using the geometry of numbers. And yet, the resulting estimates fit essentially perfectly, estimating the contribution of regular conjugacy classes to (1.16) as

$$(1.17) \quad \preccurlyeq e^{(2/p(\boldsymbol{\sigma})-1)R} \cdot e^R \Delta_k^{-1/2} \cdot \Delta_k^{1/2} |N_{k/\mathbb{Q}}(\Delta_{k(\gamma)/k})|^{1/2} \cdot |N_{k/\mathbb{Q}}(\Delta(\gamma))|^{-1/2} \cdot \frac{|N_{k/\mathbb{Q}}(\Delta(\gamma))|^{1/2}}{|N_{k/\mathbb{Q}}(\Delta_{k(\gamma)/k})|^{1/2}} w_\kappa^{\mathfrak{n}}(\mathrm{tr}\,\gamma).$$

Here $w_\kappa^{\mathfrak{n}}(\mathrm{tr}\,\gamma)$ is an explicit factor, which may be large but is $\preccurlyeq 1$ in a certain average sense (suitable for us), and $X \preccurlyeq Y$ stands for $X \ll_\varepsilon \mathcal{C}(\Gamma, \boldsymbol{T})^\varepsilon Y$. Following remarkable cancellations in (1.17), we obtain Theorem 1 by making the choice $R := 7 + \log \mathcal{C}(\Gamma, \boldsymbol{T})$, which matches the central and regular contributions.

In particular, when *counting conjugacy classes* contributing to (1.16), we first prove (see Lemma 15) that the classes in $\mathbf{G}^{\mathrm{reg}}(k)$ intersecting the union of all $\mathbf{G}(\mathbb{A}_f)$-conjugates of $K_\emptyset(\mathfrak{o})$ are determined up to $\preccurlyeq 1$ possibilities by their traces, a new feature for $\mathbf{G} = \mathrm{SL}_1(A)$. This reduces the problem to bounding the sum of $w_\kappa^{\mathfrak{n}}(x)$ over $x \in \mathfrak{o}$ (which is a weighted count of points in cosets of ideal lattices) lying in a specified, typically highly unbalanced polycylinder in $\mathbb{A}_\infty$. For this purpose, we have developed Lemma 19, a variant of a geometry of numbers result from [20], capturing that ideal lattices in $\mathbb{A}_\infty$ are not too skew, even with a varying k; see §2.4.

The emergence of $\mathcal{C}(\Gamma, \boldsymbol{T})$ in Theorem 1, as the natural guiding parameter for the density hypothesis, appears to be novel. It is essentially the *analytic conductor*, encapsulating the complexity at all places of k. Its role is best understood by rewriting (1.15) in the form in which it arises in our arguments, namely as

$$\sum_{\pi \in \mathcal{B}(\boldsymbol{\sigma}, \boldsymbol{T})} \mathrm{m}(\pi, \Gamma) \cdot e^{(1-2/p(\boldsymbol{\sigma}))R} \preccurlyeq \mathcal{C}(\Gamma, \boldsymbol{T}),$$

where $R := 7 + \log \mathcal{C}(\Gamma, \boldsymbol{T})$ is related to the allowable support of f in (1.16). The right-hand side, which includes a majorizer for the Plancherel volume of the unit ball around any $\pi \in \mathcal{B}(\boldsymbol{\sigma}, \boldsymbol{T})$ and thus arises naturally from the central terms in the trace formula detecting such π in $L^2(\Gamma \backslash G)$, is a famous barrier in the multiplicity problem that has never been overcome on a power scale. Our ability to do so for $p(\boldsymbol{\sigma}) > 2$, and in fact uniformly in both factors in (1.14), is due to our ability to increase R symmetrically in the self-normalized bound (1.17). See also [27] for estimates with power savings at all places of $k = \mathbb{Q}$ on congruence surfaces.

As in [9, Prop. 7.3], our estimates on the geometric terms (Propositions 3–5) along with Theorem 1 and some additional care with archimedean test functions (extending Proposition 2) should allow for a uniform sharp-cutoff Weyl's law for $\mathrm{m}(\pi_{i\boldsymbol{T}}, \Gamma)$ with $\boldsymbol{T}$ in rather general bounded regions in $\mathbb{R}^{S_\infty^G}$, uniformly over $\Gamma \in \mathcal{F}_{a,b,c}$ and with a $1/\log \mathcal{C}(\Gamma, \boldsymbol{T})$-savings in the error term; cf. [9, §3.1].

1.5. **Outline of the paper.** In §2, we give a detailed overview of the method, worked out in a specific instance chosen to illustrate all key steps. In §§3–5, we prove all essential ingredients to estimate contributions of non-central conjugacy classes to the trace formula. In §3, we fix the Haar measures at the archimedean places, construct the archimedean test functions, and estimate their orbital integrals. In §4, we fix the Haar measures at the non-archimedean places, and estimate the orbital integrals of characteristic functions of the relevant congruence subgroups of $\mathrm{SL}_2(\mathfrak{o}_\mathfrak{p})$. In §5.1, we adapt Borel's volume formula to our situation. In §5.2, we estimate the volume of $\mathbf{G}_\gamma(k)\backslash\mathbf{G}_\gamma(\mathbb{A})$ for regular semisimple elements $\gamma \in \mathbf{G}(k)$. In §5.3, we bound the number of rational conjugacy classes $[\gamma] \subset \mathbf{G}(k)$ of a fixed trace that bring nonzero contributions to the trace formula. Finally, in §6 we set up the global test function and combine all the ingredients to prove Theorem 1.

1.6. **Notations.** We denote by $\mathbb{N} := \{0, 1, 2, \dots\}$ the set of natural numbers. Let $a, b, c \in \mathbb{N}$ as before. Throughout this article, k is a number field of degree $a + 2b + c$, with ring of integers $\mathfrak{o}$ and ring of adeles $\mathbb{A} = \mathbb{A}_\infty \times \mathbb{A}_f$. We enumerate the archimedean places of k by $j \in \{1, \dots, a + b + c\}$ in such a way that $j \in \{a + 1, \dots, a + b\}$ correspond to all complex places of k. We write $\rho_j := [k_j : \mathbb{R}]$. For any nonzero prime ideal $\mathfrak{p} \subset \mathfrak{o}$, we write $k_\mathfrak{p}$ for the $\mathfrak{p}$-adic completion of k, $\mathfrak{o}_\mathfrak{p}$ for the ring of integers of $k_\mathfrak{p}$, and q for the size of residue field $\mathfrak{o}/\mathfrak{p}$. We also write $\widehat{\mathfrak{o}} := \prod_\mathfrak{p} \mathfrak{o}_\mathfrak{p}$. We shall use the Haar probability measure on $\widehat{\mathfrak{o}}$ and on the factors $\mathfrak{o}_\mathfrak{p}$. We write Δ_k and $N = N_{k/\mathbb{Q}}$ for the absolute discriminant and norm in $k/\mathbb{Q}$, and we also write $\Delta_{l/k}$ and $N_{l/k}$ for the relative discriminant and norm in an extension l/k (both valued in ideals in $\mathfrak{o}$), with analogous notations for local fields.

We denote by A a division quaternion algebra over k of signature (a, b, c), by $\mathbf{G}$ the algebraic group $\mathrm{SL}_1(A)$ defined over k, and by $\mathrm{ram}(A)$ the set of places ramified in A. If $\gamma \in A$, we write $k(\gamma)$ for the subfield generated by γ, which is quadratic unless γ is in the center of A. We write $[\gamma]$ for the conjugacy class of γ, in the group $\mathbf{G}(k)$ or in one of its discrete subgroups Γ, depending on the context. We denote by $\mathbf{G}^{\mathrm{reg}}(k)$ the set of regular semisimple elements of $\mathbf{G}(k)$, and by $\mathbf{G}_\gamma$ the centralizer of γ in $\mathbf{G}$, with analogous notations G^{reg} and G_γ in the group G as in the next paragraph.

In §3, G is either $\mathrm{SL}_2(\mathbb{R})$ or $\mathrm{SL}_2(\mathbb{C})$. In §4, G stands for $\mathrm{SL}_2(F)$ for a non-archimedean local field F. Outside of these two sections, we will write $G = \mathrm{SL}_2(\mathbb{R})^a \times \mathrm{SL}_2(\mathbb{C})^b$. We write $\widehat{G}$ for the unitary dual of G and $\widehat{G}_{\mathrm{sph}}$ for the subset of spherical representations.

For $x \in \mathbb{R}$, we write $x^+ := \max(x, 0)$. The notation $X \ll_{d,e,\dots} Y$ means that there exists a constant $C = C(a, b, c, d, e, \dots)$ such that $|X| \leqslant CY$. In particular, all implicit constants are allowed to depend on the signature (a, b, c) or equivalently on the degree $[k : \mathbb{Q}]$, but are otherwise absolute except as indicated by a subscript. We also write $X \asymp Y$ to indicate that $X \ll Y \ll X$. Finally, by $X \preccurlyeq Y$ we mean that $X \ll_\varepsilon \mathcal{C}(\Gamma, \boldsymbol{T})^\varepsilon Y$ holds for all $\varepsilon > 0$.

1.7. **Acknowledgements.** We are truly grateful to Gergő Nemes for proving various bounds involving the Legendre function. These bounds were not necessary in the end, but they played an important role at an earlier stage of the manuscript. We thank Tobias Finis for useful discussions about the limit multiplicity property. Djordje Milićević would like to thank the Max Planck Institute for Mathematics for their support and exceptional research infrastructure. Finally, we are grateful to the referee for valuable comments that helped us improve the exposition.

2. Method

In this section, we give a detailed outline of the proof, using a specific example to simplify matters while capturing the most important features. In this spirit, we consider the case when $G = \mathrm{SL}_2(\mathbb{R})^2$, k is a totally real field of degree 4, A is the unique quaternion algebra over k ramified exactly at the third and fourth archimedean place (i.e., $a = 2$, $b = 0$, $c = 2$), and $\Gamma = \Gamma_\emptyset(\mathfrak{o})$. We settle for bounding the multiplicity in $L^2(\Gamma\backslash G)$ of a single non-tempered spherical representation, $\pi_{(\sigma, iT)}$ for $\sigma \in (0, 1/2)$ and $T \in \mathbb{R}$, as in (1.2).

Borel's volume formula (Proposition 4) yields $\mathrm{vol}(\Gamma\backslash G) \asymp \Delta_k^{3/2}$, so we need to prove

$$(2.1) \qquad \mathrm{m}(\pi_{(\sigma, iT)}, \Gamma) \ll_\varepsilon \left(\Delta_k^{3/2}(1 + |T|) \right)^{1 - 2\sigma + \varepsilon}.$$

2.1. **Trace formula setup.** Following the approach of [11, 53] we estimate the multiplicity by the trace of a suitably chosen positive operator acting on $L^2(\Gamma\backslash G)$. Let $f \in C_c(G)$, and let $Rf : L^2(\Gamma\backslash G) \to L^2(\Gamma\backslash G)$ be

the operator

$$(2.2) \qquad (Rf\phi)(h) := \int_G \phi(hg)f(g)\,dg.$$

The quotient $\Gamma\backslash G$ is compact, so Rf is of trace class, and we have

$$(2.3) \qquad \operatorname{tr} Rf = \sum_{\pi\in\widehat{G}} \mathrm{m}(\pi,\Gamma)\operatorname{tr}\pi(f) = \sum_{[\gamma]\subset\Gamma} \operatorname{vol}(\Gamma_\gamma\backslash G_\gamma)\mathbf{O}(\gamma,f),$$

where $[\gamma]$ runs through the set of conjugacy classes of Γ, and

$$\mathbf{O}(\gamma,f) := \int_{G_\gamma\backslash G} f(g^{-1}\gamma g)\,d\dot{g}.$$

We shall choose a test function of the form $f = f_1 \otimes f_2$, where $f_1, f_2 \in C_c(\mathrm{SL}_2(\mathbb{R}))$ are strongly positive definite (cf. beginning of §1.4). We want to find f such that $\operatorname{tr}\pi_{(\sigma,iT)}(f)$ is large, while the right-hand (geometric) side of the trace formula remains relatively small. Since $f \in C_c(G)$ is strongly positive definite, this will automatically lead to an upper bound on $\mathrm{m}(\pi_{(\sigma,iT)},\Gamma)$.

We switch to the adelic version of the trace formula in order to make the orbital integrals, the volumes, and the set of conjugacy classes more manageable. Let $f_{\mathbb{A}} := f \otimes \mathbf{1}^2_{\mathrm{SU}_2(\mathbb{C})} \otimes \mathbf{1}_{K_\emptyset(\mathfrak{o})}$. In §5, we normalize the Haar measure on $\mathbf{G}(\mathbb{A})$ such that $\operatorname{vol}(\mathbf{G}(k)\backslash\mathbf{G}(\mathbb{A})) = \operatorname{vol}(\Gamma\backslash G)$. By (6.8) we have

$$(2.4) \qquad \operatorname{tr} Rf = \operatorname{tr} Rf_{\mathbb{A}} = \sum_{[\gamma]\subset\mathbf{G}(k)} \operatorname{vol}(\mathbf{G}_\gamma(k)\backslash\mathbf{G}_\gamma(\mathbb{A}))\mathbf{O}(\gamma,f_{\mathbb{A}}),$$

where the sum is now taken over the conjugacy classes of $\mathbf{G}(k)$.

Since the test function $f_{\mathbb{A}}$ is given as a pure tensor, the orbital integrals in (2.4) decompose as

$$(2.5) \qquad \mathbf{O}(\gamma,f_{\mathbb{A}}) = \mathbf{O}(\gamma,f_1)\mathbf{O}(\gamma,f_2)\prod_{v\in\mathrm{ram}_\infty(A)}\mathbf{O}(\gamma,\mathbf{1}_{\mathrm{SU}_2(\mathbb{C})})\prod_{\mathfrak{p}}\mathbf{O}(\gamma,\mathbf{1}_{\mathrm{SL}_2(\mathfrak{o}_{\mathfrak{p}})}).$$

The bounds for the local orbital integrals will naturally involve the *Weyl discriminant*, which is an element of k defined for $\gamma \in A(k) \subset \mathrm{SL}_2(\overline{k})$ with eigenvalues $w, w^{-1} \in \overline{k}^\times$ as

$$\Delta(\gamma) := (w - w^{-1})^2 = \operatorname{tr}(\gamma)^2 - 4.$$

We note that $\Delta(\gamma)$ is the negative of the more standard Weyl discriminant $\det(1 - \mathrm{Ad}(\gamma))|_{\mathfrak{g}/\mathfrak{g}_\gamma}$.

2.2. Archimedean test functions.
Let $R \geqslant 0$ be a parameter. Using Harish-Chandra's spherical transform, we construct in Propositions 1 and 2 spherical test functions $f_1, f_2 \in C_c(\mathrm{SL}_2(\mathbb{R}))$ with the following properties:

- Large trace: $\operatorname{tr}\pi_\sigma^{\mathbb{R}}(f_1) \gg e^{2\sigma R}$ and $\operatorname{tr}\pi_{iT}^{\mathbb{R}}(f_2) \gg 1$.
- Controlled support: $\operatorname{supp} f_1 \subset B(2R+2)$ and $\operatorname{supp} f_2 \subset B(2)$, where $B(R)$ is as in (3.19).
- Small orbital integrals: $\mathbf{O}(\gamma,f_j) \ll |\Delta(\gamma)|_j^{-1/2}$ for any $\gamma \in \mathbf{G}^{\mathrm{reg}}(k)$.
- Boundedness: $f_1(g) \ll 1$ and $f_2(g) \ll 1 + |T|$.

These properties imply that

- $\operatorname{tr}\pi_{(\sigma,iT)}(f) \gg e^{2\sigma R}$;
- the sum in (2.4) is taken over $[\pm\mathrm{id}]$ and the set $W(R)$ of $[\gamma] \subset \mathbf{G}^{\mathrm{reg}}(k)$ such that the conjugacy class of γ in $\mathbf{G}(\mathbb{A})$ intersects $B(2R+2) \times B(2) \times \mathrm{SU}_2(\mathbb{C})^2 \times \mathrm{SL}_2(\widehat{\mathfrak{o}})$;
- $\mathbf{O}(\gamma,f_1)\mathbf{O}(\gamma,f_2) \ll |N_{k/\mathbb{Q}}(\Delta(\gamma))|^{-1/2}$;
- $f(\pm\mathrm{id}) \ll 1 + |T|$.

The test function $f = f_1 \otimes f_2 \in C_c(G)$ is strongly positive definite, so from (2.3), (2.4) and (2.5) we get

$$(2.6) \qquad \mathrm{m}(\pi_{(\sigma,iT)}) \ll e^{-2R\sigma}\left(\operatorname{vol}(\mathbf{G}(k)\backslash\mathbf{G}(\mathbb{A}))(1+|T|) + \sum_{[\gamma]\in W(R)} \frac{\operatorname{vol}(\mathbf{G}_\gamma(k)\backslash\mathbf{G}_\gamma(\mathbb{A}))\mathbf{O}(\gamma,\mathbf{1}_{\mathrm{SL}_2(\widehat{\mathfrak{o}})})}{|N_{k/\mathbb{Q}}(\Delta(\gamma))|^{1/2}}\right).$$

The first summand comes from the central conjugacy classes $[\gamma] = [\pm \mathrm{id}]$. Our strategy is to choose the maximal $R \geqslant 0$ for which the central terms still dominate the sum (up to a negligible factor). We will see later that the correct choice is provided by

$$(2.7) \qquad e^R = e^7 \cdot \mathcal{C}(\Gamma, \boldsymbol{T}) \asymp \mathrm{vol}(\mathbf{G}(k)\backslash\mathbf{G}(\mathbb{A}))(1 + |T|) \asymp \Delta_k^{3/2}(1 + |T|).$$

The factor e^7 ensures that $R \geqslant 0$ (see Proposition 4).

2.3. **Estimating the geometric terms.**
The non-archimedean orbital integrals are estimated in §4 using Bruhat–Tits trees. This treatment allows us to achieve great uniformity in the number field k, ramification of the quaternion algebra A, and the compact open subgroup of $\mathbf{G}(\mathbb{A}_f)$ giving rise to the lattice Γ. In the present case, our Proposition 3 shows that

$$\left|\mathbf{O}(\gamma, \mathbf{1}_{\mathrm{SL}_2(\mathfrak{o}_\mathfrak{p})})\right| \leqslant \begin{cases} 1, & \text{if } |\Delta(\gamma)|_\mathfrak{p} = 1, \\ C|\Delta(\gamma)|_\mathfrak{p}^{-1/2}|\Delta_{k(\gamma)/k}|_\mathfrak{p}^{1/2}, & \text{otherwise}, \end{cases}$$

where $C > 0$ is an absolute constant. Multiplying this bound over all non-archimedean places $\mathfrak{p}$, and denoting by $\omega(\Delta(\gamma))$ the number of distinct prime ideals diving $\Delta(\gamma)$, we get

$$\left|\mathbf{O}(\gamma, \mathbf{1}_{\mathrm{SL}_2(\widehat{\mathfrak{o}})})\right| \leqslant C^{\omega(\Delta(\gamma))} \prod_\mathfrak{p} |\Delta(\gamma)|_\mathfrak{p}^{-1/2}|\Delta_{k(\gamma)/k}|_\mathfrak{p}^{1/2} \ll_\varepsilon e^{\varepsilon R} \frac{|N_{k/\mathbb{Q}}(\Delta(\gamma))|^{1/2}}{|N_{k/\mathbb{Q}}(\Delta_{k(\gamma)/k})|^{1/2}},$$

since $[\gamma] \in W(R)$ forces an upper bound $|N_{k/\mathbb{Q}}(\Delta(\gamma))| \ll e^{2R}$ (cf. (6.18)), so that $C^{\omega(\Delta(\gamma))} \ll_\varepsilon e^{\varepsilon R}$.

Volumes of adelic quotients of algebraic tori such as $\mathbf{G}_\gamma$ in (2.6) have been computed in [54]. We show in Proposition 5 that

$$\mathrm{vol}(\mathbf{G}_\gamma(k)\backslash\mathbf{G}_\gamma(\mathbb{A})) \ll_\varepsilon \Delta_k^{1/2+\varepsilon}|N_{k/\mathbb{Q}}(\Delta_{k(\gamma)/k})|^{1/2+\varepsilon}.$$

Combining the above bounds on volumes of adelic quotients and non-archimedean orbital integrals, the contribution of a single regular conjugacy class $[\gamma] \in W(R)$ to (2.6) satisfies

$$\frac{\mathrm{vol}(\mathbf{G}_\gamma(k)\backslash\mathbf{G}_\gamma(\mathbb{A}))\mathbf{O}(\gamma, \mathbf{1}_{\mathrm{SL}_2(\widehat{\mathfrak{o}})})}{|N_{k/\mathbb{Q}}(\Delta(\gamma))|^{1/2}} \ll_\varepsilon e^{4\varepsilon R}\Delta_k^{1/2}.$$

Returning to (2.6), and using also (2.7), we conclude that

$$\mathrm{m}(\pi_{(\sigma, iT)}, \Gamma) \ll_\varepsilon e^{-2R\sigma + 4\varepsilon R}\left(e^R + \Delta_k^{1/2}\#W(R)\right).$$

2.4. **Counting the contributing conjugacy classes.**
By Lemmata 15–17, the conjugacy classes $[\gamma] \in W(R)$ are "almost determined" by their traces, with no more than $\ll_\varepsilon e^{\varepsilon R}$ classes in $W(R)$ of any given trace. On the other hand, $[\gamma] \in W(R)$ implies (cf. (6.18) and the comment under (6.10)) that $\mathrm{tr}(\gamma) \in \mathfrak{o}$, $|\mathrm{tr}(\gamma)|_1 < e^{R+2}$, and $|\mathrm{tr}(\gamma)|_j < e^2$ for $j \in \{2, 3, 4\}$. Therefore,

$$\#W(R) \ll_\varepsilon e^{\varepsilon R} \cdot \#\left\{x \in \mathfrak{o} : |x|_1 < e^{R+2} \text{ and } |x|_j < e^2 \text{ for } j \in \{2, 3, 4\}\right\}.$$

Since the covolume of $\mathfrak{o}$ in $\mathbb{A}_\infty$ is $\Delta_k^{1/2}$, following the volume–covolume principle, one would expect that the count on the right-hand side is about $e^R/\Delta_k^{1/2}$. By [20, Cor. 1], this is indeed the case as long as $e^R \gg \Delta_k$, reflecting the fact that the lattice $\mathfrak{o}$ is not too skew. Therefore, substituting this upper bound on $\#W(R)$ into the above bound for $\mathrm{m}(\pi_{(\sigma, iT)}, \Gamma)$, and recalling also (2.7),

$$\mathrm{m}(\pi_{(\sigma, iT)}, \Gamma) \ll_\varepsilon e^{-2R\sigma + R + 5\varepsilon R} \ll_\varepsilon \left(\Delta_k^{3/2}(1 + |T|)\right)^{1-2\sigma + 5\varepsilon}.$$

This completes our sketch of the proof of (2.1).

3. Archimedean aspects

3.1. **Haar measures.** We specify a Haar measure on $G = \mathrm{SL}_2(\mathbb{R})$ in terms of the Iwasawa decomposition $G = ANK$. As usual, A is the subgroup of positive diagonal matrices, N is the subgroup of upper triangular unipotent matrices, and K equals $\mathrm{SO}_2(\mathbb{R})$. We define Haar measures on A and N by the formulae

$$\int_A f(a)\, da := \int_{\mathbb{R}} f(a(t))\, dt, \qquad a(t) := \begin{pmatrix} e^{t/2} & \\ & e^{-t/2} \end{pmatrix};$$

$$\int_N f(n)\, dn := \int_{\mathbb{R}} f(n(x))\, dx, \qquad n(x) := \begin{pmatrix} 1 & x \\ & 1 \end{pmatrix}.$$

We write dk for the Haar probability measure on K, and then we put

$$(3.1) \qquad \int_G f(g)\, dg := \int_K \int_N \int_A f(ank)\, da\, dn\, dk.$$

Then, [13, Th. 11.2.1] along with the normalization under [13, Th. 11.1.3] shows that the same measure in terms of the Cartan decomposition reads

$$(3.2) \qquad \int_G f(g)\, dg = 2\pi \int_K \int_0^\infty \int_K f(k_1 a(t) k_2)\, (\sinh t)\, dk_1\, dt\, dk_2.$$

On $G = \mathrm{SL}_2(\mathbb{C})$ we specify the Haar measure similarly. We start from the Iwasawa decomposition $G = ANK$, where the meaning of A and N is as before, and K equals $\mathrm{SU}_2(\mathbb{C})$. We use the same Haar measure on A as before, on N we take

$$\int_N f(n)\, dn := \int_{\mathbb{R}^2} f(n(x + iy))\, dx\, dy,$$

with $n(\cdot)$ as before, and on K we take the Haar probability measure. We define dg by (3.1), and then [35, Th. 1.7.1] shows that

$$(3.3) \qquad \int_G f(g)\, dg = 4\pi \int_K \int_0^\infty \int_K f(k_1 a(t) k_2)\, (\sinh t)^2\, dk_1\, dt\, dk_2.$$

3.2. **Orbital integrals.** Let $G = \mathrm{SL}_2(\mathbb{K})$, where $\mathbb{K}$ is either $\mathbb{R}$ or $\mathbb{C}$. We shall write ρ for the degree $[\mathbb{K} : \mathbb{R}]$, and $|\cdot|_{\mathbb{K}} = |\cdot|^\rho$ for the module function of $\mathbb{K}$ (using the terminology of Weil [56, Ch. I, §2]). Let D be the subgroup of diagonal matrices, and $T := D \cap K$ be its maximal compact subgroup. Based on the polar decomposition $D = AT$, we specify a Haar measure on D as the product of the Haar measure da on A and the Haar probability measure on T.

Let $\gamma \in G^{\mathrm{reg}}$ be a regular semisimple element with distinct eigenvalues $w, w^{-1} \in \mathbb{C}^\times$. The centralizer of γ in G, denoted by G_γ, is conjugate to D or K (but not both). More precisely, if $w \in \mathbb{K}$, then there are precisely two ways to map G_γ to D by an inner automorphism of G, and these are connected by the inverse map on D (conjugation by the Weyl element). Otherwise $\mathbb{K} = \mathbb{R}$ and $|w| = 1$, and there is a unique way to map G_γ to K by an inner automorphism of G. Depending on the case, we use a conjugation to transport the Haar measure on D or K to G_γ. As G_γ is closed, and both G and G_γ are unimodular, the coset space $G_\gamma \backslash G$ carries a unique right G-invariant measure $d\dot{g}$ such that

$$\int_G f(g)\, dg = \int_{G_\gamma \backslash G} \left(\int_{G_\gamma} f(hg)\, dh \right) d\dot{g}, \qquad f \in C_c(G).$$

Here $\dot{g}$ abbreviates the coset $G_\gamma g$. We shall think of $g \in G$ as running through a set of representatives, and write dg instead of $d\dot{g}$ for convenience.

We introduce the *orbital integral* of a compactly supported function f over the conjugacy class of γ:

$$(3.4) \qquad \mathbf{O}(\gamma, f) := \int_{G_\gamma \backslash G} f(g^{-1} \gamma g)\, dg, \qquad \gamma \in G^{\mathrm{reg}}, \quad f \in C_c(G).$$

The orbital integral has an important conjugation invariance property, which is plausible but which we spell out for clarity.

Lemma 1. *Let $\gamma, \delta \in G^{\mathrm{reg}}$ be regular semisimple elements. If γ and δ are conjugate in G, then*

$$\mathbf{O}(\gamma, f) = \mathbf{O}(\delta, f), \qquad f \in C_c(G).$$

Proof. By assumption, there is an inner automorphism $g \mapsto hgh^{-1}$ of G that maps γ to δ, and G_γ to G_δ. This automorphism transports the measure on $G_\gamma \backslash G$ to the measure on $G_\delta \backslash G$, hence

$$
\begin{aligned}
\mathbf{O}(\gamma, f) &= \int_{G_\gamma \backslash G} f(g^{-1}\gamma g) \, dg = \int_{G_\gamma \backslash G} f(g^{-1}h^{-1}\delta h g) \, dg \\
&= \int_{G_\gamma \backslash G} f(h^{-1}(hg^{-1}h^{-1})\delta(hgh^{-1})h) \, dg \\
&= \int_{G_\delta \backslash G} f(h^{-1}g^{-1}\delta gh) \, dg = \int_{G_\delta \backslash G} f(g^{-1}\delta g) \, dg = \mathbf{O}(\delta, f).
\end{aligned}
$$

In the last step, we used that the measure on $G_\delta \backslash G$ is right G-invariant. The proof is complete. $\qquad\square$

Remark 2. Conjugation by the Weyl element acts by the inverse map on D, hence Lemma 1 implies that

$$(3.5) \qquad \mathbf{O}(\gamma, f) = \mathbf{O}(\gamma^{-1}, f), \qquad \gamma \in D^{\mathrm{reg}}, \quad f \in C_c(G).$$

We shall need to estimate $\mathbf{O}(\gamma, f)$ for a given spherical function $f \in C_c(K \backslash G / K)$ and varying γ. As we do this, thanks to Lemma 1, we will be able to assume that either $\gamma \in D$, or $G = \mathrm{SL}_2(\mathbb{R})$ and $\gamma \in \mathrm{SO}_2(\mathbb{R})$. In both cases, we shall express $\mathbf{O}(\gamma, f)$ in terms of the *Harish-Chandra transform*

$$(3.6) \qquad \mathbf{H}(y, f) := |y|_{\mathbb{K}}^{1/2} \int_N f(a(\log y)n) \, dn, \qquad y > 0, \quad f \in C_c(K \backslash G / K),$$

and the *Weyl discriminant*

$$\Delta(\gamma) := (w - w^{-1})^2 = \mathrm{tr}(\gamma)^2 - 4, \qquad \gamma \in G.$$

The usage of the y-coordinate in (3.6) is justified by the fact that, by our conventions, $d(a(\log y)) = d(\log y) = dy/y$, which is also the measure used in the definition of the Mellin transform.

Lemma 2. *Let $\gamma \in D^{\mathrm{reg}}$ be a regular diagonal element with distinct eigenvalues $w, w^{-1} \in \mathbb{C}^\times$. Then*

$$|\Delta(\gamma)|_{\mathbb{K}}^{1/2} \, \mathbf{O}(\gamma, f) = \mathbf{H}(|w|^2, f), \qquad f \in C_c(K \backslash G / K).$$

Proof. Recalling the definition of the measures on G, D, $D \backslash G$, in particular recalling that on $T = D \cap K$ we use the Haar probability measure, we see that (cf. (3.1) and (3.4))

$$\mathbf{O}(\gamma, f) = \int_{D \backslash G} f(g^{-1}\gamma g) \, dg = \int_K \int_N f(k^{-1}n^{-1}\gamma n k) \, dn \, dk = \int_N f(n^{-1}\gamma n) \, dn.$$

By (3.5), we can assume without loss of generality that $\gamma = \mathrm{diag}(w, w^{-1})$. Using also the notation $n = n(z)$, a small calculation gives that the commutator $\gamma^{-1}n^{-1}\gamma n$ equals $n(z - w^{-2}z)$. Therefore, a change of variables yields that

$$\mathbf{O}(\gamma, f) = \int_N f(\gamma\gamma^{-1}n^{-1}\gamma n) \, dn = \frac{1}{|1 - w^{-2}|_{\mathbb{K}}} \int_N f(\gamma n) \, dn.$$

As $f \in C_c(K \backslash G / K)$, we can replace γ by $\mathrm{diag}(|w|, |w|^{-1})$ on the right-hand side, and hence

$$|w - w^{-1}|_{\mathbb{K}} \, \mathbf{O}(\gamma, f) = |w|_{\mathbb{K}} \int_N f(\mathrm{diag}(|w|, |w|^{-1})n) \, dn = \mathbf{H}(|w|^2, f).$$

As $|w - w^{-1}|_{\mathbb{K}}$ equals $|\Delta(\gamma)|_{\mathbb{K}}^{1/2}$, we are done. $\qquad\square$

Remark 3. Incidentally, Lemma 2 implies the following symmetry of the Harish-Chandra transform:

$$(3.7) \qquad \mathbf{H}(y, f) = \mathbf{H}(y^{-1}, f), \qquad y > 0, \quad f \in C_c(K \backslash G / K).$$

There is a nice analogue of Lemma 2 for elliptic orbital integrals. Before stating this result, we introduce some notation. For every $g \in G$, we define the *height* $H(g)$ as the unique nonnegative number such that

$$g = k_1 a(H(g)) k_2$$

for some $k_1, k_2 \in K$. That is,

$$(3.8) \qquad \cosh(H(g)) = \frac{1}{2} \operatorname{tr}(gg^*), \qquad g \in G.$$

We shall also write $1 + 2h(g)$ for the left-hand side, that is,

$$(3.9) \qquad h(g) := \frac{1}{4} \operatorname{tr}(gg^*) - \frac{1}{2}, \qquad g \in G.$$

It is clear that $h(g) \geqslant 0$, since $\cosh(H(g)) \geqslant 1$. The rationale behind these definitions is that a spherical function $f(g)$ can be thought of as a function of $H(g)$ or $h(g)$, whichever is more convenient.

The following identity for elliptic orbital integrals was inspired by a calculation in [31, §10.6], namely by [31, (10.28)] and the surrounding two displays.

Lemma 3. *Assume that* $\mathbb{K} = \mathbb{R}$. *Let* $\gamma \in \mathrm{SO}_2(\mathbb{R})$ *be a regular element with distinct eigenvalues* $e^{\pm i\theta}$ *on the unit circle. Then*

$$(3.10) \qquad |\Delta(\gamma)|_{\mathbb{K}}^{1/2} \mathbf{O}(\gamma, f) = \int_{\mathbb{R}} \frac{\mathbf{H}(e^{r(v)}, f)}{1 + v^2} \, dv, \qquad f \in C_c(K \backslash G / K),$$

where $r(v)$ *abbreviates* $2 \sinh^{-1}(v \sin \theta)$.

Proof. Using the definition of the measures on G, K, $K \backslash G$, we see that (cf. (3.2) and (3.4))

$$
\begin{aligned}
\mathbf{O}(\gamma, f) &= \int_{K \backslash G} f(g^{-1} \gamma g) \, dg \\
&= 2\pi \int_0^\infty \int_K f(k^{-1} a(-t) \gamma a(t) k) \, (\sinh t) \, dk \, dt \\
&= 2\pi \int_0^\infty f(a(-t) \gamma a(t)) \, (\sinh t) \, dt.
\end{aligned}
$$

We also observe that

$$a(-t)\gamma a(t) = \begin{pmatrix} e^{-t/2} & \\ & e^{t/2} \end{pmatrix} \begin{pmatrix} \cos\theta & \pm \sin\theta \\ \mp \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} e^{t/2} & \\ & e^{-t/2} \end{pmatrix} = \begin{pmatrix} \cos\theta & \pm e^{-t} \sin\theta \\ \mp e^t \sin\theta & \cos\theta \end{pmatrix},$$

whence by (3.9),

$$h(a(-t)\gamma a(t)) = \frac{e^{2t} + e^{-2t}}{4} \sin^2\theta + \frac{\cos^2\theta - 1}{2} = (\sinh t)^2 (\sin\theta)^2.$$

Therefore, if $F \colon [0, \infty) \to \mathbb{C}$ denotes the unique function satisfying $f(g) = F(h(g))$, then

$$(3.11) \qquad |\Delta(\gamma)|_{\mathbb{K}}^{1/2} \mathbf{O}(\gamma, f) = 4\pi |\sin\theta| \int_0^\infty F((\sinh t)^2 (\sin\theta)^2) \, (\sinh t) \, dt.$$

We need to show that the right-hand sides of (3.10) and (3.11) are equal. We start by rewriting $\mathbf{H}(e^{r(v)}, f)$ in terms of F. In general, for an arbitrary $r \in \mathbb{R}$, we have by (3.9)

$$h(a(r)n(x)) = h\left(\begin{pmatrix} e^{r/2} & e^{r/2} x \\ & e^{-r/2} \end{pmatrix} \right) = \frac{e^r x^2}{4} + \sinh^2 \frac{r}{2},$$

hence by (3.6) also

$$\mathbf{H}(e^r, f) = e^{r/2} \int_{\mathbb{R}} F\left(\frac{e^r x^2}{4} + \sinh^2 \frac{r}{2} \right) dx = 2|\sin\theta| \int_{\mathbb{R}} F\left(x^2 \sin^2\theta + \sinh^2 \frac{r}{2} \right) dx.$$

We apply this to $r(v) := 2\sinh^{-1}(v\sin\theta)$ in place of r, and integrate the resulting expression against the measure $dv/(1+v^2)$:

$$\int_{\mathbb{R}} \frac{\mathbf{H}(e^{r(v)}, f)}{1+v^2}\, dv = 2|\sin\theta| \int_{\mathbb{R}} \int_{\mathbb{R}} \frac{F(x^2\sin^2\theta + v^2\sin^2\theta)}{1+v^2}\, dx\, dv$$

$$= 8|\sin\theta| \int_0^\infty \int_0^\infty \frac{F(x^2\sin^2\theta + v^2\sin^2\theta)}{1+v^2}\, dx\, dv.$$

We make the polar-like change of variables

$$x = (\sinh t)(\cos\phi) \qquad \text{and} \qquad v = (\sinh t)(\sin\phi),$$

where $t > 0$ and $0 < \phi < \pi/2$. The map $(t, \phi) \mapsto (x, y)$ is a diffeomorphism from $(0, \infty) \times (0, \pi/2)$ to $(0, \infty)^2$ with absolute Jacobian determinant $(\sinh t)(\cosh t)$. Therefore,

$$\int_{\mathbb{R}} \frac{\mathbf{H}(e^{r(v)}, f)}{1+v^2}\, dv = 8|\sin\theta| \int_0^\infty \int_0^{\pi/2} \frac{F((\sinh t)^2(\sin\theta)^2)}{1 + (\sinh t)^2(\sin\phi)^2} (\sinh t)(\cosh t)\, d\phi\, dt.$$

However, for every $t \in \mathbb{R}$,

$$\int_0^{\pi/2} \frac{\cosh t}{1 + (\sinh t)^2(\sin\phi)^2}\, d\phi = \left[\tan^{-1}\left((\cosh t)(\tan\phi)\right) \right]_0^{\pi/2-} = \frac{\pi}{2}.$$

As a result,

$$\int_{\mathbb{R}} \frac{\mathbf{H}(e^{r(v)}, f)}{1+v^2}\, dv = 4\pi|\sin\theta| \int_0^\infty F((\sinh t)^2(\sin\theta)^2)\,(\sinh t)\, dt.$$

The right-hand side is identical to the right-hand side of (3.11), hence we are done. $\qquad\square$

Remark 4. It follows from Lemma 1 combined with (3.5) and (3.11) that

$$\mathbf{O}(\gamma, f) = \mathbf{O}(\gamma^{-1}, f), \qquad \gamma \in G^{\mathrm{reg}}, \quad f \in C_c(K\backslash G/K).$$

We shall use Lemmata 1–3 to estimate the relevant archimedean orbital integrals by the corresponding Harish-Chandra transforms.

Lemma 4. *Let $\gamma \in G^{\mathrm{reg}}$ be a regular semisimple element. Then*

$$|\Delta(\gamma)|_{\mathbb{K}}^{1/2} |\mathbf{O}(\gamma, f)| \leqslant \pi \sup_{y>0} |\mathbf{H}(y, f)|, \qquad f \in C_c(K\backslash G/K).$$

Proof. By Lemma 1, we can assume that either $\gamma \in D$, or $G = \mathrm{SL}_2(\mathbb{R})$ and $\gamma \in \mathrm{SO}_2(\mathbb{R})$. In the first case, the bound follows trivially from Lemma 2. In the second case, the bound follows from Lemma 3 combined with the triangle inequality for integrals; the constant π is the integral of $1/(1+v^2)$ over $\mathbb{R}$. $\qquad\square$

We end this subsection with an elementary estimate to be used in §6.3.

Lemma 5. *Let $\gamma \in G$. Then $|\mathrm{tr}(\gamma)|^2$ and $|\Delta(\gamma)|$ are at most $4\cosh(H(\gamma))$.*

Proof. Let $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in G$, so that $\mathrm{tr}(\gamma) = a + d$ and $\Delta(\gamma) = (a+d)^2 - 4 = (a-d)^2 + 4bc$. Then $|\mathrm{tr}(\gamma)|^2$ and $|\Delta(\gamma)|$ are at most $2(|a|^2 + |b|^2 + |c|^2 + |d|^2)$, which in turn equals $4\cosh(H(\gamma))$ by (3.8). $\qquad\square$

3.3. **Spherical functions.** For $s \in \mathbb{C}$, we denote by $\phi_s \in C^\infty(K\backslash G/K)$ the elementary spherical function of normalized Casimir eigenvalue $1/4 - s^2$. In particular,

$$(3.12) \qquad\qquad \phi_{-s}(g) = \phi_s(g) = \phi_s(g^{-1}).$$

The function $\phi_s(g)$ can be realized as the matrix coefficient $\langle \pi_s(g)v, v \rangle$, where (π_s, V_s) is the spherical principal series representation of G unitarily induced from the character $\mathrm{diag}(y^{1/2}, y^{-1/2}) \mapsto |y|_{\mathbb{K}}^s$, and $v \in V_s$ is a K-invariant unit vector. In this language, the imaginary axis $s \in i\mathbb{R}$ corresponds to the tempered spectrum, while the real interval $s \in (-1/2, 1/2)$ corresponds to the exceptional spectrum (complementary series). In particular,

$$(3.13) \qquad\qquad |\phi_s(g)| \leqslant 1, \qquad s \in [-1/2, 1/2] \cup i\mathbb{R}.$$

In terms of special functions, we have

$$\phi_s(g) = P_{-1/2+s}(\cosh(H(g))), \qquad \mathbb{K} = \mathbb{R}; \tag{3.14}$$

$$\phi_s(g) = \frac{\sinh(2sH(g))}{2s\sinh(H(g))}, \qquad \mathbb{K} = \mathbb{C}. \tag{3.15}$$

Here, $P_{-1/2+s}$ denotes the Legendre function, and $H(g)$ is given by (3.8). See [31, p. 23], [39, pp. 47, 84, 202], [14, (22.6.11.7)], [25, p. 433, (30)], [21, §§1.5–1.8], [35, Prop. 2.4.5] for more details.

The *spherical transform* of $f \in C_c^\infty(K\backslash G/K)$ is given by

$$\widehat{f}(s) := \int_G f(g)\,\phi_s(g)\,dg, \qquad s \in \mathbb{C}. \tag{3.16}$$

Thus $\widehat{f}(s) = \langle \pi_s(f)v, v\rangle$ for a K-invariant unit vector $v \in V_s$ as above. It follows that $\pi_s(f)$ acts by the scalar $\widehat{f}(s)$ on V_s^K, and by zero on the orthogonal complement $V_s^{K,\perp}$. In particular,

$$\widehat{f*g}(s) = \widehat{f}(s)\widehat{g}(s), \qquad f, g \in C_c^\infty(K\backslash G/K), \quad s \in \mathbb{C}. \tag{3.17}$$

The spherical transform can also be understood as the *normalized Mellin transform* of the Harish-Chandra transform (cf. [39, Ch. 5, Th. 5] and [35, Th. 1.4.1]):

$$\widehat{f}(s) = \int_0^\infty |y|_\mathbb{K}^s\,\mathbf{H}(y, f)\,\frac{dy}{y} = \int_0^\infty y^{\rho s}\,\mathbf{H}(y, f)\,\frac{dy}{y}. \tag{3.18}$$

Based on this connection, the spherical transform $f \mapsto \widehat{f}$ provides an isomorphism between the convolution algebra $C_c^\infty(K\backslash G/K)$ and the usual algebra of even entire functions satisfying a Paley–Wiener type condition. More precisely, let us introduce the notation

$$B(R) := \{g \in G: \ H(g) \leqslant R\}, \qquad R > 0. \tag{3.19}$$

Using the known inverses of the Mellin and Harish-Chandra transforms (cf. [39, Ch. 5, Th. 3–5]), it is straightforward to verify that the following three conditions are equivalent:

- f is supported in $B(R)$;
- $\mathbf{H}(y, f)$ is supported in $[e^{-R}, e^R]$;
- $\widehat{f}(\sigma + i\tau) \ll_N e^{\rho R|\sigma|}(1 + |\tau|)^{-N}$ for every $\sigma, \tau \in \mathbb{R}$ and $N > 0$.

The inverse spherical transform on the even Paley–Wiener space can be given directly as

$$f(g) = \frac{1}{2\pi}\int_\mathbb{R} \phi_{i\tau}(g)\,\widehat{f}(i\tau)\,\tau\tanh(\pi\tau)\,d\tau, \qquad \mathbb{K} = \mathbb{R}; \tag{3.20}$$

$$f(g) = \frac{2}{\pi^2}\int_\mathbb{R} \phi_{i\tau}(g)\,\widehat{f}(i\tau)\,\tau^2\,d\tau, \qquad \mathbb{K} = \mathbb{C}. \tag{3.21}$$

The leading constants here are rather ad-hoc as they depend on the normalization of the Haar measure on G. In fact (3.20) is equivalent to the inversion formula for the classical Mehler–Fock transform, as can be seen from (3.2), (3.14), (3.16). Similarly, (3.21) is equivalent to the inversion formula for the classical sine transform, as can be seen from (3.3), (3.15), (3.16). See [39, Ch. V], [35, Ch. 1], [21, §§1.9–1.10] for more details. We note that [39, Ch. V, Th. 6–7] are slightly in error, e.g. $\mathbf{S}f(i\tau)$ should be $\mathbf{S}f(2i\tau)$, which is apparently our $\widehat{f}(i\tau)$.

In the previous section, we saw how to estimate an orbital integral in terms of the Harish-Chandra transform. Now we can state and prove a similar bound in terms of the spherical transform.

Lemma 6. *Let $\gamma \in G^{\mathrm{reg}}$ be a regular semisimple element. Then*

$$|\Delta(\gamma)|_\mathbb{K}^{1/2}\,|\mathbf{O}(\gamma, f)| \leqslant \int_\mathbb{R} |\widehat{f}(i\tau)|\,d\tau, \qquad f \in C_c(K\backslash G/K).$$

Proof. By (3.18) we have, for any $y > 0$,

$$\mathbf{H}(y, f) = \frac{1}{2\pi i}\int_{i\mathbb{R}} y^{-s}\,\widehat{f}\left(\frac{s}{\rho}\right)\,ds = \frac{\rho}{2\pi}\int_\mathbb{R} y^{-i\tau\rho}\,\widehat{f}(i\tau)\,d\tau.$$

Hence the result follows from Lemma 4 and the triangle inequality for integrals. □

14

3.4. **Construction of test functions.** Let us call a function $F \in C_c(G)$ *strongly positive definite* if it is a convolution $u * \breve{u}$, where $u \in C_c(G)$ and

$$(3.22) \qquad \breve{u}(g) := \overline{u(g^{-1})}, \qquad g \in G.$$

Then F is positive definite in the usual sense [15, §13.4.1 & §13.4.11], and $\pi(F) = \pi(u)\pi(u)^*$ is a positive operator for every $\pi \in \widehat{G}$.

Recall that $G = \mathrm{SL}_2(\mathbb{K})$, where $\mathbb{K}$ is either $\mathbb{R}$ or $\mathbb{C}$, and $\rho = [\mathbb{K} : \mathbb{R}]$. We shall construct a strongly positive definite test function $F \in C_c^\infty(K \backslash G / K)$ with well-controlled support and size, such that the spherical transform $\widehat{F}(s)$ is sufficiently large for certain spectral parameters $s \in \mathbb{C}$, while the orbital integrals $\mathbf{O}(\gamma, F)$ are sufficiently small for all regular semisimple elements $\gamma \in G^{\mathrm{reg}}$. These estimates will be crucial in our application of the trace formula. As in the original construction of Sarnak–Xue [53, (28)–(29)], we shall provide F in the form of $u * \breve{u}$, where $u, \breve{u} \in C_c^\infty(K \backslash G / K)$ are connected by (3.22). In terms of the spherical transform, this means that (cf. (3.12), (3.16), (3.17))

$$(3.23) \qquad \widehat{F}(s) = \widehat{u}(s)\overline{\widehat{u}(\overline{s})}, \qquad s \in \mathbb{C}.$$

We fix, once and for all, a smooth function $f \colon K \backslash G / K \to [0, \infty)$, which is supported in $B(1/2)$ but is not identically zero. Then, the Harish-Chandra transform $\mathbf{H}(y, f)$ is nonnegative for every $y > 0$, and it is supported in $[e^{-1/2}, e^{1/2}]$. Using (3.7) and (3.18), we see that

$$\widehat{f}(s) = \int_0^\infty \cosh(\rho s \log y) \, \mathbf{H}(y, f) \, \frac{dy}{y},$$

hence $\widehat{f}(s)$ is real for $s \in \mathbb{R} \cup i\mathbb{R}$, and

$$(3.24) \qquad \widehat{f}(s) \geqslant \cos(1) \, \widehat{f}(0) > 0, \qquad s \in [-1, 1] \cup i[-1, 1].$$

Moreover,

$$(3.25) \qquad \widehat{f}(\sigma + i\tau) \ll_N e^{\rho|\sigma|/2}(1 + |\tau|)^{-N}, \qquad \sigma, \tau \in \mathbb{R}, \quad N > 0.$$

We shall define $F(g)$ in terms of $f(g)$. Our explicit constructions appear in the proofs of the two propositions below, which will be used at non-tempered and tempered archimedean places, respectively.

Proposition 1. *Let $R \geqslant 0$. There exists $u \in C_c^\infty(K \backslash G / K)$ such that the strongly positive definite spherical function $F := u * \breve{u}$ has the following properties:*

- $\widehat{F}(\sigma) \gg e^{2\rho R|\sigma|}$ *for every $\sigma \in [-1, 1]$;*
- *F is supported in $B(2R + 2)$, and it satisfies the bound $F \ll 1$;*
- $|\Delta(\gamma)|_{\mathbb{K}}^{1/2} \, |\mathbf{O}(\gamma, F)| \ll 1$ *for every $\gamma \in G^{\mathrm{reg}}$.*

The implied constants are absolute.

Proof. Let us define $u \in C_c^\infty(K \backslash G / K)$ through its spherical transform

$$(3.26) \qquad \widehat{u}(s) := \cosh(\rho R s) \, \widehat{f}(s)^2, \qquad s \in \mathbb{C}.$$

This function is even, entire, and by (3.25) it satisfies the Paley–Wiener type condition

$$\widehat{u}(\sigma + i\tau) \ll_N e^{\rho(R+1)|\sigma|}(1 + |\tau|)^{-N}, \qquad \sigma, \tau \in \mathbb{R}, \quad N > 0.$$

Hence indeed $\widehat{u}(s)$ is the spherical transform of a unique $u \in C_c^\infty(K \backslash G / K)$, which is supported in $B(R + 1)$. By (3.23) and a similar reasoning (or by direct calculation), it is clear now that $F := u * \breve{u}$ is supported in $B(2R + 2)$. We record the following simple consequence of (3.23) and (3.26):

$$\widehat{F}(s) = \cosh(\rho R s)^2 \, \widehat{f}(s)^4, \qquad s \in \mathbb{C}.$$

Combining this with (3.24), we obtain for every $\sigma \in [-1, 1]$ that

$$\widehat{F}(\sigma) \geqslant \cosh(\rho R \sigma)^2 \, \cos(1)^4 \, \widehat{f}(0)^4 \gg e^{2\rho R|\sigma|}.$$

Similarly, using the spherical inversion formulae (3.20)–(3.21) and the upper bounds (3.13) and (3.25), we obtain for every $g \in G$ that

$$F(g) \leqslant \frac{2}{\pi^2} \int_{\mathbb{R}} \widehat{f}(i\tau)^4 \, |\tau|^\rho \, d\tau \ll 1.$$

15

Finally, by (3.25) and Lemma 6, we infer for every $\gamma \in G^{\mathrm{reg}}$ that

$$|\Delta(\gamma)|_{\mathbb{K}}^{1/2}\,|\mathbf{O}(\gamma,F)| \leqslant \int_{\mathbb{R}} \widehat{f}(i\tau)^4\,d\tau \ll 1.$$

The proof is complete. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

Proposition 2. *Let $t \in \mathbb{R}$. There exists $u \in C_c^\infty(K\backslash G/K)$ such that the strongly positive definite spherical function $F := u * \breve{u}$ has the following properties:*

- $\widehat{F}(i\tau) \gg 1$ *for every* $\tau \in [t-1, t+1]$;
- F *is supported in* $B(2)$, *and it satisfies the bound* $F \ll (1+|t|)^\rho$;
- $|\Delta(\gamma)|_{\mathbb{K}}^{1/2}\,|\mathbf{O}(\gamma,F)| \ll 1$ *for every* $\gamma \in G^{\mathrm{reg}}$.

The implied constants are absolute.

Proof. Let us define $u \in C_c^\infty(K\backslash G/K)$ through its spherical transform

(3.27) $$\widehat{u}(s) := \widehat{f}(s-it)^2 + \widehat{f}(s+it)^2, \qquad s \in \mathbb{C}.$$

This function is even, entire, and by (3.25) it satisfies the Paley–Wiener type condition

$$\widehat{u}(\sigma + i\tau) \ll_{t,N} e^{\rho|\sigma|}(1+|\tau|)^{-N}, \qquad \sigma, \tau \in \mathbb{R}, \quad N > 0.$$

Hence indeed $\widehat{u}(s)$ is the spherical transform of a unique $u \in C_c^\infty(K\backslash G/K)$, which is supported in $B(1)$. By (3.23) and a similar reasoning (or by direct calculation), it is clear now that $F := u * \breve{u}$ is supported in $B(2)$. We record the following simple consequence of (3.23) and (3.27):

$$\widehat{F}(s) = \left(\widehat{f}(s-it)^2 + \widehat{f}(s+it)^2\right)^2, \qquad s \in \mathbb{C}.$$

Combining this with (3.24), we obtain for every $\tau \in [t-1, t+1]$ that

$$\widehat{F}(i\tau) \geqslant \cos(1)^4\,\widehat{f}(0)^4 \gg 1.$$

Similarly, using the spherical inversion formulae (3.20)–(3.21) and the upper bounds (3.13) and (3.25), we obtain for every $g \in G$ that

$$F(g) \leqslant \frac{8}{\pi^2} \int_{\mathbb{R}} \widehat{f}(i\tau - it)^4\,|\tau|^\rho\,d\tau \ll (1+|t|)^\rho.$$

In the last step, we used that $|\tau|^\rho$ is less than the product of $(1+|\tau-t|)^\rho$ and $(1+|t|)^\rho$. Finally, by (3.25) and Lemma 6, we infer for every $\gamma \in G^{\mathrm{reg}}$ that

$$|\Delta(\gamma)|_{\mathbb{K}}^{1/2}\,|\mathbf{O}(\gamma,F)| \leqslant 4 \int_{\mathbb{R}} \widehat{f}(i\tau)^4\,d\tau \ll 1.$$

The proof is complete. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

4. Non-archimedean aspects

4.1. **Haar measures and orbital integrals.** Let F be a non-archimedean local field with ring of integers $\mathfrak{o}$, maximal ideal $\mathfrak{p} \subset \mathfrak{o}$, and uniformizer $\varpi \in \mathfrak{p}$ (which we fix throughout). As usual, we define $v_\mathfrak{p}\colon \mathfrak{o} \to \mathbb{N} \cup \{\infty\}$ by setting $v_\mathfrak{p}(x) = r$ for $x \in \varpi^r \mathfrak{o}^\times$ and $v_\mathfrak{p}(0) = \infty$. We write $q := \#(\mathfrak{o}/\mathfrak{p})$ for the size of the residue field, and we normalize the multiplicative valuation $|\cdot|_F$ on F so that $|\varpi|_F = q^{-1}$. We set

$$G := \mathrm{SL}_2(F) \qquad \text{and} \qquad K := \mathrm{SL}_2(\mathfrak{o}),$$

and we normalize the Haar measure on G so that K has measure 1.

A regular semisimple element $\gamma \in G^{\mathrm{reg}}$ has two distinct eigenvalues $w, w^{-1} \in \overline{F}^\times$, and we shall distinguish between the cases $w \in F$ (γ is split) and $w \notin F$ (γ is elliptic). In both cases, we write $E := F(w)$ for the splitting field of γ, and we fix some $\tau \in \mathrm{GL}_2(E)$ such that $\tau^{-1}\gamma\tau = \mathrm{diag}(w, w^{-1})$. In the elliptic case, we choose τ specifically as follows. We start from a nonzero vector $\left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right) \in E^2$ satisfying $\gamma\left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right) = w\left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right)$. Then, $\gamma\left(\begin{smallmatrix} \overline{x} \\ \overline{y} \end{smallmatrix}\right) = \overline{w}\left(\begin{smallmatrix} \overline{x} \\ \overline{y} \end{smallmatrix}\right)$, where the upper bar indicates the action of the nontrivial element of $\mathrm{Gal}(E/F)$. On the other hand, $\overline{w}$ equals w^{-1}, hence $\tau := \left(\begin{smallmatrix} x & \overline{x} \\ y & \overline{y} \end{smallmatrix}\right)$ has the required property. We note that this τ does not lie in $\mathrm{SL}_2(E)$.

We specify a Haar measure on the centralizer $G_\gamma \leqslant G$ of γ as follows. If γ is split, then
$$G_\gamma = \tau \left\{ \operatorname{diag}(a, a^{-1}) : a \in F^\times \right\} \tau^{-1},$$
which has an inner direct product decomposition $G_\gamma = \Lambda_\gamma T_\gamma$ with
$$\Lambda_\gamma := \tau \left\{ \operatorname{diag}(\varpi^n, \varpi^{-n}) : n \in \mathbb{Z} \right\} \tau^{-1},$$
$$T_\gamma := \tau \left\{ \operatorname{diag}(a, a^{-1}) : a \in \mathfrak{o}^\times \right\} \tau^{-1}.$$
We observe that the lattice $\Lambda_\gamma \leqslant G_\gamma$ and the compact open subgroup $T_\gamma \leqslant G_\gamma$ are uniquely determined by γ. On Λ_γ we take the counting measure, on T_γ we take the Haar probability measure, and on G_γ we take the product of these two measures. If γ is elliptic, then
$$G_\gamma = G \cap \tau \left\{ \operatorname{diag}(a, a^{-1}) : a \in E^\times \right\} \tau^{-1}.$$
That is, G_γ consists of the matrices $g = \tau \cdot \operatorname{diag}(a, a^{-1}) \cdot \tau^{-1}$ such that $a \in E^\times$ and $\bar{g} = g$. Using the relation $\bar{\tau} = \tau \left(\begin{smallmatrix} & 1 \\ 1 & \end{smallmatrix} \right)$, a quick calculation reveals that $\bar{g} = g$ is equivalent to $a\bar{a} = 1$. On the other hand, a multiplicative valuation of E is invariant under $\operatorname{Gal}(E/F)$, hence $a\bar{a} = 1$ can only hold when a is a unit in E. We conclude that G_γ has the transparent description
$$G_\gamma = \tau \left\{ \operatorname{diag}(a, \bar{a}) : a \in \mathfrak{o}_E^\times \text{ and } a\bar{a} = 1 \right\} \tau^{-1}.$$
In particular, G_γ is a compact group, and we take the Haar probability measure on it.

Now we can repeat what we did in the archimedean situation. As G_γ is closed, and both G and G_γ are unimodular, the coset space $G_\gamma \backslash G$ carries a unique right G-invariant measure $d\dot{g}$ such that
$$\int_G f(g)\, dg = \int_{G_\gamma \backslash G} \left(\int_{G_\gamma} f(hg)\, dh \right) d\dot{g}, \qquad f \in C_c(G).$$
Here $\dot{g}$ abbreviates the coset $G_\gamma g$. We shall think of $g \in G$ as running through a set of representatives, and write dg instead of $d\dot{g}$ for convenience. When γ is split, we specify a right G-invariant measure on $\Lambda_\gamma \backslash G$ in the same way, using the already chosen measures on Λ_γ and G.

We introduce the *orbital integral* of a compactly supported function f over the conjugacy class of γ:
$$\mathbf{O}(\gamma, f) := \int_{G_\gamma \backslash G} f(g^{-1}\gamma g)\, dg, \qquad \gamma \in G^{\mathrm{reg}}, \quad f \in C_c(G).$$

Just as in the case of archimedean local fields, we have the following.

Lemma 7. *Let $\gamma, \delta \in G^{\mathrm{reg}}$ be regular semisimple elements. If γ and δ are conjugate in G, then*
$$\mathbf{O}(\gamma, f) = \mathbf{O}(\delta, f), \qquad f \in C_c(G).$$

Proof. Same as that of Lemma 1. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad\square$

The next lemma computes the orbital integral of the characteristic function of a compact open subgroup $U \leqslant G$. With later applications in mind, it is convenient to formulate it in a more general form which also includes a normalizer of U.

Lemma 8. *Let $U \leqslant V \leqslant G$ be two compact open subgroups such that V normalizes U. Let $\gamma \in G^{\mathrm{reg}}$ be a regular semisimple element. If γ is split, then*
$$\mathbf{O}(\gamma, \mathbf{1}_U) = \operatorname{vol}(V) \cdot \# \left\{ g \in \Lambda_\gamma \backslash G / V : \gamma g U = g U \right\}.$$
If γ is elliptic, then
$$\mathbf{O}(\gamma, \mathbf{1}_U) = \operatorname{vol}(V) \cdot \# \left\{ g \in G / V : \gamma g U = g U \right\}.$$

Proof. In both cases, the result follows by a simple calculation using the left G_γ-invariance and the right V-invariance of the function
$$f(g) := \mathbf{1}_U(g^{-1}\gamma g), \qquad g \in G,$$
along with the definition of the measures on G_γ and $G_\gamma \backslash G$ (as well as Λ_γ and $\Lambda_\gamma \backslash G$, if γ is split). If $\gamma \in G^{\mathrm{reg}}$ is split, then
$$\mathbf{O}(\gamma, \mathbf{1}_U) = \int_{G_\gamma \backslash G} f(g)\, dg = \int_{\Lambda_\gamma \backslash G} f(g)\, dg = \operatorname{vol}(V) \sum_{g \in \Lambda_\gamma \backslash G / V} \mathbf{1}_{\gamma g U = g U}.$$

17

The last equation can be checked by noting that if g runs through a set of representatives for $\Lambda_\gamma \backslash G/V$, the disjoint union of the corresponding left cosets gV form a set of representatives for $\Lambda_\gamma \backslash G$.

If $\gamma \in G^{\mathrm{reg}}$ is elliptic, then

$$\mathbf{O}(\gamma, \mathbf{1}_U) = \int_{G_\gamma \backslash G} f(g)\, dg = \int_G f(g)\, dg = \mathrm{vol}(V) \sum_{g \in G/V} \mathbf{1}_{\gamma g U = g U}.$$

The proof is complete. $\qquad \square$

4.2. **Congruence subgroups and the main estimate.** For a nonzero ideal $\mathfrak{n} \subset \mathfrak{o}$, we introduce the congruence subgroups

$$(4.1) \qquad K_0(\mathfrak{n}) := \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in K : c \in \mathfrak{n} \right\},$$

$$(4.2) \qquad K_1(\mathfrak{n}) := \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in K : a - d, b, c \in \mathfrak{n} \right\}.$$

Note that $K_0(\mathfrak{n})$ and $K_1(\mathfrak{n})$ are (closely related to) the local counterparts of the classical congruence subgroups $\Gamma_0(n)$ and $\Gamma(n)$ (not $\Gamma_1(n)$).

The next two results are well-known to experts, but for convenience we provide short proofs.

Lemma 9. *For a nonzero ideal $\mathfrak{n} \subset \mathfrak{o}$, we have*

$$K_1(\mathfrak{n}) = \bigcap_{k \in K} k^{-1} K_0(\mathfrak{n}) k.$$

Proof. $K_1(\mathfrak{n})$ is a normal subgroup of K, because it is the kernel of the reduction map $K \to \mathrm{PSL}_2(\mathfrak{o}/\mathfrak{n})$. Therefore, the left-hand side is contained in the right-hand side. Now let $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ be an element of the right-hand side. Then kgk^{-1} lies in $K_0(\mathfrak{n})$ for every $k \in K$. Applying this for the matrices $\begin{pmatrix} 1 & \\ & 1 \end{pmatrix}$, $\begin{pmatrix} & -1 \\ 1 & \end{pmatrix}$, $\begin{pmatrix} 1 & \\ 1 & 1 \end{pmatrix}$ in the role of k, we obtain that the lower left entries c, $-b$, $a - b + c - d$ lie in $\mathfrak{n}$. Hence $g \in K_1(\mathfrak{n})$, and we are done. $\qquad \square$

Lemma 10. *Let r be a positive integer, and let t be the size of the 2-torsion of $(\mathfrak{o}/\mathfrak{p}^r)^\times$. Then*

$$[K : K_0(\mathfrak{p}^r)] = q^r(1 + q^{-1}) \qquad and \qquad [K : K_1(\mathfrak{p}^r)] = q^{3r}(1 - q^{-2})/t.$$

Proof. First, we verify the identity for $[K : K_0(\mathfrak{p}^r)]$. By an induction argument, it suffices to show that

$$(4.3) \qquad [K_0(\mathfrak{p}^{r-1}) : K_0(\mathfrak{p}^r)] = \begin{cases} q + 1, & \text{if } r = 1, \\ q, & \text{if } r \geqslant 2. \end{cases}$$

Two matrices $\begin{pmatrix} a & b \\ c & d \end{pmatrix}, \begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix} \in K$ lie in the same left coset of $K_0(\mathfrak{p}^r)$ if and only if $ac' - a'c \in \mathfrak{p}^r$. Using this relation, one obtains a set of representatives for $K_0(\mathfrak{p}^{r-1})/K_0(\mathfrak{p}^r)$ by picking the matrices $\begin{pmatrix} 1 & \\ c & 1 \end{pmatrix}$, where c runs through a set of representatives for $\mathfrak{p}^{r-1}/\mathfrak{p}^r$, and also picking the additional matrix $\begin{pmatrix} & -1 \\ 1 & \end{pmatrix}$ when $r = 1$. This proves (4.3).

Now, we verify the identity for $[K : K_1(\mathfrak{p}^r)]$. The quotient group $K_0(\mathfrak{p}^r)/K_1(\mathfrak{p}^r)$ is isomorphic to the group of upper triangular matrices in $\mathrm{PSL}_2(\mathfrak{o}/\mathfrak{p}^r)$. The latter group has cardinality $\#(\mathfrak{o}/\mathfrak{p}^r) \cdot \#(\mathfrak{o}/\mathfrak{p}^r)^\times/t$, hence by the already established identity for $[K : K_0(\mathfrak{p}^r)]$,

$$[K : K_1(\mathfrak{p}^r)] = [K : K_0(\mathfrak{p}^r)] \cdot [K_0(\mathfrak{p}^r) : K_1(\mathfrak{p}^r)] = q^r(1 + q^{-1}) \cdot q^{2r}(1 - q^{-1})/t.$$

The proof is complete. $\qquad \square$

As in the archimedean case, we introduce the *Weyl discriminant*

$$\Delta(\gamma) := (w - w^{-1})^2 = \mathrm{tr}(\gamma)^2 - 4, \qquad \gamma \in G.$$

It will be convenient to use the notation

$$\nu(\gamma) := v_{\mathfrak{p}}(\Delta(\gamma))/2, \qquad \gamma \in G^{\mathrm{reg}}.$$

Let $K^{\mathrm{reg}} := K \cap G^{\mathrm{reg}}$, and define, for $r \in \mathbb{N}$ and $j \in \{0,1\}$, the functions $w_j^r \colon K^{\mathrm{reg}} \to \mathbb{R}_{\geqslant 0}$ as

$$(4.4) \qquad w_0^r(\gamma) := \begin{cases} q^{\min(\nu(\gamma), \lfloor r/2 \rfloor)}, & \text{if } \gamma \text{ is split,} \\ q^{\lfloor r/2 \rfloor} \mathbf{1}_{\mathfrak{p}^r | \Delta(\gamma)}, & \text{if } \gamma \text{ is elliptic;} \end{cases}$$

$$(4.5) \qquad w_1^r(\gamma) := q^{2r} \mathbf{1}_{\mathfrak{p}^{2r} | \Delta(\gamma)}.$$

The rest of this section is devoted to the proof of the following estimate on non-archimedean orbital integrals.

Proposition 3. *For $\gamma \in K^{\mathrm{reg}}$, $r \in \mathbb{N}$, $j \in \{0,1\}$, we have*

$$(4.6) \qquad |\Delta(\gamma)|_F^{1/2} \, \mathbf{O}(\gamma, \mathbf{1}_{K_j(\mathfrak{p}^r)}) \ll q^{-\lambda/2} \operatorname{vol}(K_j(\mathfrak{p}^r)) w_j^r(\gamma),$$

where $\lambda = 0$ when E/F is unramified, and $\lambda = 1$ when E/F is ramified. The implied constant is absolute. Moreover, if $|\Delta(\gamma)|_F = 1$, then

$$(4.7) \qquad \mathbf{O}(\gamma, \mathbf{1}_{K_j(\mathfrak{p}^r)}) \leqslant \operatorname{vol}(K_j(\mathfrak{p}^r)).$$

4.3. **The Bruhat–Tits tree.** An important tool to reduce the proof of Proposition 3 to a graph theoretic enumeration is the Bruhat–Tits tree, of which we recall some important properties from [38, §5]. The *Bruhat–Tits tree* is a graph $\mathcal{X}_F$ with vertex set $\mathrm{PGL}_2(F)/\mathrm{PGL}_2(\mathfrak{o})$. Two vertices $u, v \in \mathcal{X}_F$ are connected by an edge if and only if they can be represented by $g, h \in \mathrm{GL}_2(F)$ such that $g^{-1}h = \operatorname{diag}(\varpi, 1)$. Then $\mathcal{X}_F$ is a $(q+1)$-regular tree, which we shall also regard as a simplicial complex with vertices and edges being the zero- and one-dimensional simplices, respectively. We shall write $[u, v]$ for the unique path joining two vertices $u, v \in \mathcal{X}_F$.

We equip $\mathcal{X}_F$ with the geodesic metric by assigning length 1 to each edge, and we denote by $\operatorname{dist}_F(U, V)$ the distance of two non-empty subsets $U, V \subset \mathcal{X}_F$. The group $\mathrm{GL}_2(F)$ acts on the left by isometries; the action is transitive on the vertex set. The action of $G = \mathrm{SL}_2(F)$ decomposes the vertex set into two orbits, with no edges between vertices from the same orbit. Infinite paths are called *apartments*, and since $\mathrm{GL}_2(F)$ also acts transitively on the set of apartments, we now select a particular one as follows. The *standard apartment* $\mathcal{A}_0$ is the infinite path with vertices $v_n := \operatorname{diag}(\varpi^n, 1)\mathrm{PGL}_2(\mathfrak{o})$ for $n \in \mathbb{Z}$. We record that $\{v_n : n \in \mathbb{Z} \text{ even}\}$ lies in one G-orbit, and $\{v_n : n \in \mathbb{Z} \text{ odd}\}$ lies in the other G-orbit.

A finite Galois extension E/F admits a similar Bruhat–Tits tree $\mathcal{X}_E$ with vertex set $\mathrm{PGL}_2(E)/\mathrm{PGL}_2(\mathfrak{o}_E)$. The Galois group $\operatorname{Gal}(E/F)$ acts on $\mathcal{X}_E$ by isometries, and if we multiply the distances in $\mathcal{X}_F$ by the ramification degree of E/F, then $\mathcal{X}_F$ gets contained in the set of fixed points as a metric subspace: $\mathcal{X}_F \subset \mathcal{X}_E^{\operatorname{Gal}(E/F)}$. By a rather deep result of Rousseau [49, Ch. V], we have equality here unless E/F is wildly ramified. A shorter proof was given by Prasad [48], but it still relies on a lot of background theory, including earlier chapters of Rousseau's unpublished thesis. These references deal with general Bruhat–Tits buildings and valued fields. For our special situation we can give a quick, direct proof.

Lemma 11. *Let E/F be a finite Galois extension. If E/F is unramified or tamely ramified, then*

$$\mathcal{X}_F = \mathcal{X}_E^{\operatorname{Gal}(E/F)}.$$

Proof. We need to show that every point $v \in \mathcal{X}_E^{\operatorname{Gal}(E/F)}$ lies in $\mathcal{X}_F$. We claim that we can restrict to the vertices of $\mathcal{X}_E$. Indeed, if v is not a vertex of $\mathcal{X}_E$, then it lies on a unique edge e of $\mathcal{X}_E$. Either e lies in $\mathcal{X}_F$, or the unique path from v to $\mathcal{X}_F$ starts with a segment of e. In both cases, we see that $e \subset \mathcal{X}_E^{\operatorname{Gal}(E/F)}$, justifying our claim.

Let $\mathfrak{P}$ denote the maximal ideal of $\mathfrak{o}_E$, and let $\varpi_E \in \mathfrak{P}$ be a uniformizer. The vertices of $\mathcal{X}_E$ can be written in the Iwasawa form

$$(4.8) \qquad v = \begin{pmatrix} \varpi_E^n & x \\ & 1 \end{pmatrix} \mathrm{PGL}_2(\mathfrak{o}_E), \qquad n \in \mathbb{Z}, \quad x \in E.$$

The integer $n \in \mathbb{Z}$ and the image of x in $E/\mathfrak{P}^n$ are uniquely determined by v, independently of the choice of ϖ_E. Indeed, a simple calculation shows that for $x, x' \in E$ we have

$$(4.9) \qquad \begin{pmatrix} \varpi_E^n & x \\ & 1 \end{pmatrix} \mathrm{PGL}_2(\mathfrak{o}_E) = \begin{pmatrix} \varpi_E^n & x' \\ & 1 \end{pmatrix} \mathrm{PGL}_2(\mathfrak{o}_E) \qquad \Longleftrightarrow \qquad x - x' \in \mathfrak{P}^n.$$

First we consider the case when E/F is unramified; we choose ϖ_E to be the uniformizer ϖ of F. We claim that every vertex $v \in \mathcal{X}_E^{\mathrm{Gal}(E/F)}$ is a vertex of $\mathcal{X}_F$. Writing v in the form (4.8), and using (4.9), we see the following: $x - x^\sigma \in \mathfrak{P}^n$ holds for all $\sigma \in \mathrm{Gal}(E/F)$, and we need to show that $x - \xi \in \mathfrak{P}^n$ holds for a suitable $\xi \in F$. The residue field $\mathfrak{o}_E/\mathfrak{P}$ is represented by the set A consisting of zero and the roots of unity in E of order coprime to q. We have a unique Teichmüller representation

$$x = \sum_{m \in \mathbb{Z}} a_m \varpi^m,$$

where $a_m \in A$, and $a_m = 0$ for m sufficiently small. The Galois group acts on x by acting on the digits a_m, hence the condition $x - x^\sigma \in \mathfrak{P}^n$ translates to $a_m = a_m^\sigma$ for all $m < n$. We conclude that $a_m \in F$ for all $m < n$, and hence

$$\xi := \sum_{m < n} a_m \varpi^m$$

is an element of F with the required property $x - \xi \in \mathfrak{P}^n$.

Now we consider the case when E/F is tamely and fully ramified; the degree $e := [E : F]$ is coprime to q. We claim that every vertex $v \in \mathcal{X}_E^{\mathrm{Gal}(E/F)}$ lies on an edge of $\mathcal{X}_F$. Writing v in the form (4.8), and combining (4.9) with the fact that the Galois translates of ϖ_E are uniformizers, we see (as in the unramified case) that $x - x^\sigma \in \mathfrak{P}^n$ holds for all $\sigma \in \mathrm{Gal}(E/F)$. The Galois average

$$\xi := \frac{1}{e} \cdot \sum_{\sigma \in \mathrm{Gal}(E/F)} x^\sigma$$

clearly lies in F, and it satisfies

$$x - \xi = \frac{1}{e} \cdot \sum_{\sigma \in \mathrm{Gal}(E/F)} (x - x^\sigma) \in \mathfrak{P}^n.$$

Therefore, using (4.9) again,

$$v = \begin{pmatrix} \varpi_E^n & \xi \\ & 1 \end{pmatrix} \mathrm{PGL}_2(\mathfrak{o}_E).$$

If n is divisible by e, then ϖ_E^n can be replaced by $\varpi^{n/e}$, and we conclude that v is a vertex of $\mathcal{X}_F$. Otherwise, v is not a vertex of $\mathcal{X}_F$, but it lies on a path of $\mathcal{X}_E$ that is identified with an edge of $\mathcal{X}_F$:

$$v \in \left[\begin{pmatrix} \varpi^{\lfloor n/e \rfloor} & \xi \\ & 1 \end{pmatrix} \mathrm{PGL}_2(\mathfrak{o}_E), \begin{pmatrix} \varpi^{\lceil n/e \rceil} & \xi \\ & 1 \end{pmatrix} \mathrm{PGL}_2(\mathfrak{o}_E) \right].$$

Finally, in the general tamely ramified case, let F_1 be the maximal unramified subextension of E/F. Then noting that F_1/F is Galois, and combining the two cases above,

$$\mathcal{X}_F = \mathcal{X}_{F_1}^{\mathrm{Gal}(F_1/F)} = \left(\mathcal{X}_E^{\mathrm{Gal}(E/F_1)} \right)^{\mathrm{Gal}(F_1/F)} = \mathcal{X}_E^{\mathrm{Gal}(E/F)}.$$

The proof is complete. $\qquad\square$

For the purpose of our counting arguments, we introduce some technical notations:

$$G_U := \{ g \in G : gu = u \text{ for all } u \in U \}, \qquad U \subset \mathcal{X}_F,$$

$$\mathcal{B}_F(U, r) := \{ x \in \mathcal{X}_F : \mathrm{dist}_F(x, U) \leqslant r \}, \qquad \emptyset \neq U \subset \mathcal{X}_F, \quad r \geqslant 0.$$

For $u \in \mathcal{X}_F$, we simply write $\mathcal{B}_F(u, r)$ for $\mathcal{B}_F(\{u\}, r)$.

Lemma 12. *For $r \in \mathbb{N}$ we have*

$$K_0(\mathfrak{p}^r) = G_{[v_{-r}, v_0]} \qquad and \qquad K_1(\mathfrak{p}^r) = G_{\mathcal{B}_F(v_0, r)}.$$

Proof. It is straightforward to check that $K = G_{\{v_0\}}$, and hence

(4.10) $$K_0(\mathfrak{p}^r) = \mathrm{diag}(\varpi^{-r}, 1) K \, \mathrm{diag}(\varpi^r, 1) \cap K = G_{\{v_{-r}, v_0\}} = G_{[v_{-r}, v_0]}.$$

Combining this result with Lemma 9, we get that

$$K_1(\mathfrak{p}^r) = \bigcap_{k \in K} G_{[kv_{-r}, v_0]} = G_U,$$

where U is the union of the paths $[kv_{-r}, v_0]$ for $k \in K$. It remains to prove that $U = \mathcal{B}_F(v_0, r)$. Equivalently, every path of length r starting at v_0 is of the form $[kv_{-r}, v_0]$ with $k \in K$. We could deduce this statement from the transitivity of $\mathrm{GL}_2(F)$ on the set of apartments, but we prefer an argument based on counting. We can assume that $r \geqslant 1$, in which case the number of paths of length r starting at v_0 equals $(q+1)q^{r-1}$. By (4.10) and Lemma 10, the number of paths $[kv_{-r}, v_0]$ with $k \in K$ also equals $(q+1)q^{r-1}$, so we are done. $\qquad\square$

The next lemma is the heart of our counting arguments. We state it in a stronger form than needed, partly for completeness, partly to underline that our counting is rather precise.

Lemma 13. *Consider the set of fixed points $\mathcal{X}_F^\gamma$ of some $\gamma \in K^{\mathrm{reg}}$ acting on $\mathcal{X}_F$.*

- (a) *If γ is split, then $\nu(\gamma) \in \mathbb{N}$ and $\mathcal{X}_F^\gamma = \mathcal{B}_F(\mathcal{A}, \nu(\gamma))$ for a suitable apartment $\mathcal{A} \subset \mathcal{X}_F$.*
- (b) *If γ is elliptic and the splitting extension E/F is unramified, then $\nu(\gamma) \in \mathbb{N}$ and $\mathcal{X}_F^\gamma = \mathcal{B}_F(v, \nu(\gamma))$ for a suitable vertex $v \in \mathcal{X}_F$.*
- (c) *If γ is elliptic and the splitting extension E/F is ramified, then $\mathcal{X}_F^\gamma \subset \mathcal{B}_F(e, \nu(\gamma) - \frac{1}{2})$ for a suitable edge $e \subset \mathcal{X}_F$, with equality in the tamely ramified case. In addition, $\nu(\gamma) \in \frac{1}{2} + \frac{1}{2}\mathbb{N}$, and $\nu(\gamma) \in \frac{1}{2} + \mathbb{N}$ in the tamely ramified case.*

Proof. (a) The eigenvalues $w, w^{-1} \in F^\times$ are units, because their sum $\mathrm{tr}(\gamma)$ is an integer. So we can assume, without loss of generality, that $\gamma = \mathrm{diag}(w, w^{-1})$ with $w \in \mathfrak{o}^\times$. Then we claim that $\mathcal{X}_F^\gamma = \mathcal{B}_F(\mathcal{A}_0, \nu(\gamma))$, where $\mathcal{A}_0$ is the standard apartment defined shortly before Lemma 11. That is,

$$\gamma v = v \qquad \Longleftrightarrow \qquad \mathrm{dist}(v, \mathcal{A}_0) \leqslant v_\mathfrak{p}(w - w^{-1}), \qquad v \in \mathcal{X}_F.$$

It suffices to show this equivalence for the vertices of $\mathcal{X}_F$. Indeed, if v is not a vertex, then it lies on a unique edge e. Either e lies in $\mathcal{A}_0$, or the unique path from v to $\mathcal{A}_0$ starts with a segment of e. In both cases, we see that $v \in \mathcal{X}_F^\gamma$ is equivalent to $e \subset \mathcal{X}_F^\gamma$, while $v \in \mathcal{B}_F(\mathcal{A}_0, \nu(\gamma))$ is equivalent to $e \subset \mathcal{B}_F(\mathcal{A}_0, \nu(\gamma))$.

The vertices of $\mathcal{X}_F$ can be written in the Iwasawa form

$$v = \begin{pmatrix} y & \\ & 1 \end{pmatrix} \begin{pmatrix} 1 & x \\ & 1 \end{pmatrix} v_0, \qquad y \in F^\times, \quad x \in F.$$

Keeping in mind that $\gamma = \mathrm{diag}(w, w^{-1}) \in \mathrm{GL}_2(\mathfrak{o})$, we infer

$$\gamma v = v \qquad \Longleftrightarrow \qquad \begin{pmatrix} 1 & -x \\ & 1 \end{pmatrix} \begin{pmatrix} w & \\ & w^{-1} \end{pmatrix} \begin{pmatrix} 1 & x \\ & 1 \end{pmatrix} \in \mathrm{GL}_2(\mathfrak{o})$$

$$\Longleftrightarrow \qquad -v_\mathfrak{p}(x) \leqslant v_\mathfrak{p}(w - w^{-1}).$$

Hence it suffices to show that

(4.11) $$\mathrm{dist}(v, \mathcal{A}_0) = \max(0, -v_\mathfrak{p}(x)).$$

For this purpose, we introduce the notation

$$\tilde{v}_n := \begin{pmatrix} y & \\ & 1 \end{pmatrix} \begin{pmatrix} 1 & x \\ & 1 \end{pmatrix} v_n = \begin{pmatrix} \varpi^n y & \\ & 1 \end{pmatrix} \begin{pmatrix} 1 & \varpi^{-n}x \\ & 1 \end{pmatrix} v_0, \qquad n \in \mathbb{Z}.$$

Then $v = \tilde{v}_0$, and we see that $\tilde{v}_n \in \mathcal{A}_0$ is equivalent to $n \leqslant v_\mathfrak{p}(x)$. Hence the unique path from v to $\mathcal{A}_0$ has vertex set

$$\{\tilde{v}_n : n \in \mathbb{Z} \text{ and } \min(0, v_\mathfrak{p}(x)) \leqslant n \leqslant 0\},$$

and (4.11) follows. The proof of part (a) is complete.

Before turning to the proof of parts (b) and (c), let us make some initial observations. The ratio of $v_\mathfrak{P}(\Delta(\gamma))$ and $v_\mathfrak{p}(\Delta(\gamma))$ equals the ramification degree of E/F. Hence from part (a), or rather its proof, it follows that

(4.12) $$\mathcal{X}_E^\gamma = \begin{cases} \mathcal{B}_E(\tau\mathcal{A}_0, \nu(\gamma)), & \text{if } E/F \text{ is unramified,} \\ \mathcal{B}_E(\tau\mathcal{A}_0, 2\nu(\gamma)), & \text{if } E/F \text{ is ramified.} \end{cases}$$

We also see for any $n \in \mathbb{Z}$ that

$$\overline{\tau v_n} = \overline{\tau} v_n = \tau \begin{pmatrix} & 1 \\ 1 & \end{pmatrix} \begin{pmatrix} \varpi^n & \\ & 1 \end{pmatrix} \mathrm{PGL}_2(\mathfrak{o}_E) = \tau \begin{pmatrix} 1 & \\ & \varpi^n \end{pmatrix} \begin{pmatrix} & 1 \\ 1 & \end{pmatrix} \mathrm{PGL}_2(\mathfrak{o}_E) = \tau v_{-n}.$$

Therefore, the generator of $\mathrm{Gal}(E/F)$ flips $\tau\mathcal{A}_0$ from end to end with a unique fixed point τv_0.

(b) Since E/F is an unramified quadratic extension,

$$\nu(\gamma) = \frac{1}{2}v_{\mathfrak{p}}(\Delta(\gamma)) = \frac{1}{2}v_{\mathfrak{P}}(\Delta(\gamma)) = v_{\mathfrak{P}}(w - \overline{w}) \in \mathbb{N}.$$

Moreover, $\mathcal{X}_E$ is a (q^2+1)-regular tree, which contains $\mathcal{X}_F$ as a spanned subtree. By Lemma 11, we see that $\tau\mathcal{A}_0 \cap \mathcal{X}_F$ consists of a single vertex v of $\mathcal{X}_F$, and every point of $\mathcal{X}_F$ is connected to $\tau\mathcal{A}_0$ through v. Using also (4.12), we obtain

$$\mathcal{X}_F^{\gamma} = \mathcal{X}_E^{\gamma} \cap \mathcal{X}_F = \mathcal{B}_E(\tau\mathcal{A}_0, \nu(\gamma)) \cap \mathcal{X}_F = \mathcal{B}_F(v, \nu(\gamma)).$$

(c) Since E/F is a ramified quadratic extension, $\mathrm{Gal}(E/F)$ acts trivially on the residue field $\mathfrak{o}_E/\mathfrak{P}$, hence

$$2\nu(\gamma) = v_{\mathfrak{p}}(\Delta(\gamma)) = \frac{1}{2}v_{\mathfrak{P}}(\Delta(\gamma)) = v_{\mathfrak{P}}(w - \overline{w}) \in 1 + \mathbb{N}.$$

Writing w as $a + b\varpi_E$ with $a, b \in \mathfrak{o}$, we see that $v_{\mathfrak{P}}(w - \overline{w})$ has the same parity as $v_{\mathfrak{P}}(\varpi_E - \overline{\varpi_E})$, therefore $2\nu(\gamma)$ is odd in the tamely ramified case. Moreover, $\mathcal{X}_E$ is a $(q+1)$-regular tree, which contains $\mathcal{X}_F$ such that every edge of $\mathcal{X}_F$ is the union of two edges of $\mathcal{X}_E$. Let $u \in \tau\mathcal{A}_0$ and $v \in \mathcal{X}_F$ be such that $\mathrm{dist}_E(u, v)$ is minimal. It is clear that u and v are vertices of $\mathcal{X}_E$ fixed by $\mathrm{Gal}(E/F)$. In particular, $u = \tau v_0$. By Lemma 11, we also have $u = v$ in the tamely ramified case. Clearly, u is not a vertex of $\mathcal{X}_F$, because its neighbors in $\tau\mathcal{A}_0$ are not fixed by $\mathrm{Gal}(E/F)$. Similarly, v is not a vertex of $\mathcal{X}_F$, either because $u = v$, or because the neighbor of v in the path $[u, v]$ lies outside $\mathcal{X}_F$. So v is the midpoint of an edge e of $\mathcal{X}_F$. Now every point $x \in \mathcal{X}_F$ is connected to $\tau\mathcal{A}_0$ through the path $[u, v]$, hence

$$\mathrm{dist}_E(x, \tau\mathcal{A}_0) = \mathrm{dist}_E(x, v) + \mathrm{dist}_E(v, u) \geqslant \mathrm{dist}_E(x, v) = 2\,\mathrm{dist}_F(x, v),$$

with equality in the tamely ramified case. Using also (4.12), we obtain

$$\mathcal{X}_F^{\gamma} = \mathcal{X}_E^{\gamma} \cap \mathcal{X}_F = \mathcal{B}_E(\tau\mathcal{A}_0, 2\nu(\gamma)) \cap \mathcal{X}_F \subset \mathcal{B}_F(v, \nu(\gamma)) = \mathcal{B}_F(e, \nu(\gamma) - 1/2),$$

with equality in the tamely ramified case.

The proof is complete in all cases. $\qquad\square$

4.4. **Proof of Proposition 3.** In this proof, by a slight abuse of notation, we shall denote by $[x, y]$ the unique *directed path* in $\mathcal{X}_F$ that starts at the vertex x and ends at the vertex y. That is, $[x, y] = [x', y']$ if and only if $x = x'$ and $y = y'$.

4.4.1. *The case of γ split.* By Lemma 7, we may assume that $\gamma \in K^{\mathrm{reg}}$ is diagonal. Then, G_{γ} consists of all diagonal matrices within G, the elements of Λ_{γ} are the matrices $\mathrm{diag}(\varpi^n, \varpi^{-n})$ for $n \in \mathbb{Z}$, and the apartment stabilized by γ is $\mathcal{A}_0$.

First we consider the case $j = 0$. By Lemma 12, we can think of the congruence subgroup $K_0(\mathfrak{p}^r)$ geometrically as $G_{[v_{-r}, v_0]}$. Hence, applying Lemma 8 to

$$U = V = K_0(\mathfrak{p}^r) = G_{[v_{-r}, v_0]},$$

we obtain

$$(4.13) \qquad \mathbf{O}(\gamma, \mathbf{1}_{K_0(\mathfrak{p}^r)}) = \mathrm{vol}(K_0(\mathfrak{p}^r)) \cdot \#\{g \in \Lambda_{\gamma}\backslash G/G_{[v_{-r}, v_0]} : \gamma g G_{[v_{-r}, v_0]} = g G_{[v_{-r}, v_0]}\}.$$

The map $g \mapsto [gv_{-r}, gv_0]$ gives rise to a bijection between $G/G_{[v_{-r}, v_0]}$ and the orbit $G \cdot [v_{-r}, v_0]$ of directed paths of length r in $\mathcal{X}_F$, with a compatible action of G on these two sets. The condition $\gamma g G_{[v_{-r}, v_0]} = g G_{[v_{-r}, v_0]}$ is equivalent to $[gv_{-r}, gv_0] \subset \mathcal{X}_F^{\gamma}$, hence by Lemma 13,

$$(4.14) \qquad \mathbf{O}(\gamma, \mathbf{1}_{K_0(\mathfrak{p}^r)}) \leqslant \mathrm{vol}(K_0(\mathfrak{p}^r)) \cdot \#(\Lambda_{\gamma}\backslash\{[x, y] \subset \mathcal{B}_F(\mathcal{A}_0, \nu(\gamma)) : \mathrm{dist}_F(x, y) = r\}).$$

This is the enumeration problem we have indicated earlier. For each directed path $[x, y] \subset \mathcal{B}_F(\mathcal{A}_0, \nu(\gamma))$, we define a vertex $v \in \mathcal{A}_0$ as follows. If $[x, y] \cap \mathcal{A}_0$ has at most one element, then v is the point of $\mathcal{A}_0$ closest to $[x, y]$. Otherwise $[x, y] \cap \mathcal{A}_0$ is a path of positive length, and v is its endpoint closer to x. The group Λ_{γ} acts on $\mathcal{A}_0$ by even shifts, hence we can assume that $v \in \{v_0, v_1\}$.

We consider two cases. Either $[x, y] \cap \mathcal{A}_0$ has at most one element, or $[x, y] \cap \mathcal{A}_0$ is a path of positive length. In the first case, let $p \in [x, y]$ be the vertex closest to $\mathcal{A}_0$. As both $\mathrm{dist}_F(v, p) + \mathrm{dist}_F(p, x)$ and $\mathrm{dist}_F(v, p) + \mathrm{dist}_F(p, y)$ are at most $\nu(\gamma)$, the distance of p from v is at most $\nu(\gamma) - r/2$; in particular, the first case is void when $r > 2\nu(\gamma)$. Recording $s := \min(\mathrm{dist}_F(p, x), \mathrm{dist}_F(p, y))$, for every $(r - \nu(\gamma))^+ \leqslant s \leqslant \lfloor r/2 \rfloor$

there are at most $6q^{\nu(\gamma)-(r-s)}$ choices for p, and for each p, there are at most $2q^r$ choices for $[x,y]$. In total, there are at most $24q^{\nu(\gamma)+\lfloor r/2 \rfloor}$ choices for $[x,y]$ in the first case.

In the second case, let w be the endpoint of the path $[x,y] \cap \mathcal{A}_0$ closer to y. We record the lengths $s := \mathrm{dist}_F(x,v)$ and $t := \mathrm{dist}_F(w,y)$; then, $[v,w]$ is a directed subpath of $[x,y]$ lying in $\mathcal{A}_0$ of length $r-s-t \geqslant 1$. Considering first without loss of generality the case $s \leqslant t$, for every choice of $0 \leqslant s \leqslant \min(\nu(\gamma), \lfloor \frac{r-1}{2} \rfloor)$ and $0 \leqslant t \leqslant \min(r-1-s, \nu(\gamma))$, there are four choices for $[v,w]$, followed by q^{s+t} choices for $[x,v]$ and $[w,y]$. In total we find for every $0 \leqslant s \leqslant \min(\nu(\gamma), \lfloor \frac{r-1}{2} \rfloor)$ at most $8q^{\min(r-1,\nu(\gamma)+s)}$ choices for $[x,y]$. Summing over s and taking into account the symmetry $s \leftrightarrow t$, we infer that there are at most

$$16\lfloor \tfrac{r+1}{2} \rfloor q^{r-1}\mathbf{1}_{1 \leqslant r \leqslant \nu(\gamma)} + 16\left(2 + \nu(\gamma) - \lceil \tfrac{r-1}{2} \rceil\right)q^{r-1}\mathbf{1}_{\nu(\gamma)+1 \leqslant r \leqslant 2\nu(\gamma)} + 32q^{2\nu(\gamma)}\mathbf{1}_{r \geqslant 2\nu(\gamma)+1}$$

choices for $[x,y]$ in the second case. Altogether, the count on the right-hand side of (4.14) is

$$\ll q^{\nu(\gamma)+\lfloor r/2 \rfloor}\mathbf{1}_{r \leqslant 2\nu(\gamma)} + q^{r-1}\min(r, 2\nu(\gamma)+1-r)\mathbf{1}_{1 \leqslant r \leqslant 2\nu(\gamma)} + q^{2\nu(\gamma)}\mathbf{1}_{r \geqslant 2\nu(\gamma)+1} \ll q^{\nu(\gamma)+\min(\nu(\gamma),\lfloor r/2 \rfloor)}.$$

This verifies (4.6) in the case of γ split and $j = 0$, upon noting that $|\Delta(\gamma)|_F^{1/2} = q^{-\nu(\gamma)}$ and $E = F$.

Now we consider the case $j = 1$. By Lemma 12, we can think of the congruence subgroup $K_1(\mathfrak{p}^r)$ geometrically as $G_{\mathcal{B}_F(v_0,r)}$. Hence, applying Lemma 8 to

$$U = K_1(\mathfrak{p}^r) = G_{\mathcal{B}_F(v_0,r)} \qquad \text{and} \qquad V = K,$$

we obtain

$$\mathbf{O}(\gamma, \mathbf{1}_{K_1(\mathfrak{p}^r)}) = \#\{g \in \Lambda_\gamma \backslash G/K : \gamma g G_{\mathcal{B}_F(v_0,r)} = g G_{\mathcal{B}_F(v_0,r)}\}.$$

As in the case of $j = 0$, the map $g \mapsto gv_0$ gives rise to a bijection between G/K and the orbit $G \cdot v_0$ of vertices in $\mathcal{X}_F$, with a compatible left action of G on these two sets. The condition $\gamma g G_{\mathcal{B}_F(v_0,r)} = g G_{\mathcal{B}_F(v_0,r)}$ is equivalent to $\mathcal{B}_F(gv_0, r) \subset \mathcal{X}_F^\gamma$, hence by Lemma 13,

$$(4.15) \qquad \mathbf{O}(\gamma, \mathbf{1}_{K_1(\mathfrak{p}^r)}) \leqslant \#(\Lambda_\gamma \backslash \{x \text{ a vertex of } \mathcal{X}_F : \mathcal{B}_F(x,r) \subset \mathcal{B}_F(\mathcal{A}_0, \nu(\gamma))\}).$$

As we divide by the action of Λ_γ, we may assume that the closest point of $\mathcal{A}_0$ to x is v_0 or v_1. Then there are at most $2q^{\nu(\gamma)-r}\mathbf{1}_{r \leqslant \nu(\gamma)}$ choices for x. Noting also

$$(4.16) \qquad \mathrm{vol}(K_1(\mathfrak{p}^r))^{-1} = [K : K_1(\mathfrak{p}^r)] < q^{3r}$$

along with $|\Delta(\gamma)|_F^{1/2} = q^{-\nu(\gamma)}$ and $E = F$, we obtain (4.6) in the case of γ split and $j = 1$.

If $|\Delta(\gamma)|_F = 1$, then $\nu(\gamma) = 0$, and we need to prove the more precise bound (4.7). For $j = 0$, similarly as in (4.13) and (4.14), we have to count those directed paths $[x,y] \subset \mathcal{A}_0$ modulo Λ_γ which belong to the G-orbit of $[v_{-r}, v_0]$. The directed paths in question form the Λ_γ-orbit of $[v_{-r}, v_0]$, hence

$$\mathbf{O}(\gamma, \mathbf{1}_{K_0(\mathfrak{p}^r)}) = \mathrm{vol}(K_0(\mathfrak{p}^r)).$$

For $j = 1$, the condition $\mathcal{B}_F(x,r) \subset \mathcal{A}_0$ forces $r = 0$, and then $K_0(\mathfrak{p}^0) = K_1(\mathfrak{p}^0)$ reduces the proof to the earlier discussed case $j = 0$.

4.4.2. *The case of γ elliptic and E/F unramified.* From Lemma 13 we know that $\mathcal{X}_F^\gamma$ equals $\mathcal{B}_F(v, \nu(\gamma))$ for a suitable vertex $v \in \mathcal{X}_F$, where also $\nu(\gamma) \in \mathbb{N}$. Then, the analogue of (4.14) reads (cf. Lemmata 8 and 12)

$$(4.17) \qquad \mathbf{O}(\gamma, \mathbf{1}_{K_0(\mathfrak{p}^r)}) \leqslant \mathrm{vol}(K_0(\mathfrak{p}^r)) \cdot \#\{[x,y] \subset \mathcal{B}_F(v, \nu(\gamma)) : \mathrm{dist}_F(x,y) = r\}.$$

For any directed path $[x,y] \subset \mathcal{B}_F(v, \nu(\gamma))$, let $p \in [x,y]$ be the vertex closest to v. As both $\mathrm{dist}_F(v,p) + \mathrm{dist}_F(p,x)$ and $\mathrm{dist}_F(v,p) + \mathrm{dist}_F(p,y)$ are at most $\nu(\gamma)$, the distance of p from v is at most $\nu(\gamma) - r/2$; so there is nothing to count when $r > 2\nu(\gamma)$. Recording $s := \min(\mathrm{dist}_F(p,x), \mathrm{dist}_F(p,y))$, for every $(r-\nu(\gamma))^+ \leqslant s \leqslant \lfloor r/2 \rfloor$ there are less than $3q^{\nu(\gamma)-(r-s)}$ choices for p, and for each p, there are at most $2q^r$ choices for $[x,y]$. Altogether, the count on the right-hand side of (4.17) is

$$\ll q^{\nu(\gamma)+\lfloor r/2 \rfloor}\mathbf{1}_{r \leqslant 2\nu(\gamma)}.$$

This verifies (4.6) in the case of γ elliptic with unramified splitting field and $j = 0$, upon noting that $|\Delta(\gamma)|_F^{1/2} = q^{-\nu(\gamma)}$ and $\lambda = 0$.

Similarly, the analogue of (4.15) reads (cf. Lemmata 8 and 12)

$$(4.18) \qquad \mathbf{O}(\gamma, \mathbf{1}_{K_1(\mathfrak{p}^r)}) \leqslant \#\{x \text{ a vertex of } \mathcal{X}_F : \mathcal{B}_F(x,r) \subset \mathcal{B}_F(v, \nu(\gamma))\}.$$

Clearly, there are less than $3q^{\nu(\gamma)-r}\mathbf{1}_{r\leqslant\nu(\gamma)}$ choices for x. Noting also (4.16) along with $|\Delta(\gamma)|_F^{1/2} = q^{-\nu(\gamma)}$ and $\lambda = 0$, we obtain (4.6) in the case of γ elliptic with unramified splitting field and $j = 1$.

If $|\Delta(\gamma)|_F = 1$, then $\nu(\gamma) = 0$, and we need to prove the more precise bound (4.7). In this case, both counts on the right-hand sides of (4.17) and (4.18) are equal to $\mathbf{1}_{r=0}$, hence (4.7) follows.

4.4.3. *The case of γ elliptic and E/F ramified.* From Lemma 13 we know that $\mathcal{X}_F^\gamma \subset \mathcal{B}_F(e, \nu(\gamma) - 1/2)$ for a suitable edge $e \subset \mathcal{X}_F$, where also $2\nu(\gamma) - 1 \in \mathbb{N}$. Hence the analogue of (4.14) and (4.17) reads (cf. Lemmata 8 and 12)

$$(4.19) \qquad \mathbf{O}(\gamma, \mathbf{1}_{K_0(\mathfrak{p}^r)}) \leqslant \mathrm{vol}(K_0(\mathfrak{p}^r)) \cdot \#\{[x,y] \subset \mathcal{B}_F(e, \nu(\gamma) - 1/2) : \mathrm{dist}_F(x,y) = r\}.$$

For estimating the number of directed paths $[x,y] \subset \mathcal{B}_F(e, \nu(\gamma) - 1/2)$ of length r, we distinguish between two cases. If $[x,y]$ contains e, then we argue similarly as in the subcase of the split case when $[x,y]$ intersected $\mathcal{A}_0$ in a path of positive length; in fact, the present case is a bit easier since the lengths of the (up to two) components of $[x,y] \setminus e$ determine each other. Otherwise, we argue as in the unramified elliptic case, with $\lfloor \nu(\gamma) - 1/2 \rfloor$ in place of $\nu(\gamma)$. Altogether, the count on the right-hand side of (4.19) is

$$\ll q^{r-1}\min(r, 2\nu(\gamma) + 1 - r)\mathbf{1}_{1\leqslant r\leqslant 2\nu(\gamma)} + q^{\lfloor\nu(\gamma)-1/2\rfloor+\lfloor r/2\rfloor}\mathbf{1}_{r\leqslant 2\nu(\gamma)-1} \ll q^{\nu(\gamma)-1/2+\lfloor r/2\rfloor}\mathbf{1}_{r\leqslant 2\nu(\gamma)}.$$

This verifies (4.6) in the case of γ elliptic with ramified splitting field and $j = 0$, upon noting that $|\Delta(\gamma)|_F^{1/2} = q^{-\nu(\gamma)}$ and $\lambda = 1$.

Similarly, the analogue of (4.15) and (4.18) reads (cf. Lemmata 8 and 12)

$$\mathbf{O}(\gamma, \mathbf{1}_{K_1(\mathfrak{p}^r)}) \leqslant \#\{x \text{ a vertex of } \mathcal{X}_F : \mathcal{B}_F(x,r) \subset \mathcal{B}_F(e, \nu(\gamma) - 1/2)\}.$$

Clearly, there are at most $4q^{\nu(\gamma)-1/2-r}\mathbf{1}_{r\leqslant\nu(\gamma)-1/2}$ choices for x. Noting also (4.16) along with $|\Delta(\gamma)|_F^{1/2} = q^{-\nu(\gamma)}$ and $\lambda = 1$, we obtain (4.6) in the case of γ elliptic with ramified splitting field and $j = 1$.

The proof of Proposition 3 is complete in all cases.

5. Global aspects

We collect some notations important for the following two sections. Let $a, b, c \in \mathbb{N}$ with $a + b \geqslant 1$, and recall that all implied constants are allowed to depend on these parameters. Let k be a number field with $a + c$ real and b complex places. Let A be a division quaternion algebra over k of signature (a, b, c). Let

$$\mathrm{ram}(A) = \mathrm{ram}_\infty(A) \cup \mathrm{ram}_f(A)$$

be the set of places (resp. archimedean places and non-archimedean places) where A ramifies. Let $\mathbf{G} := \mathrm{SL}_1(A)$ and $\mathbf{H} := \mathrm{GL}_1(A)$ viewed as algebraic groups over k, and let $G := \mathrm{SL}_2(\mathbb{R})^a \times \mathrm{SL}_2(\mathbb{C})^b$. As $\mathrm{ram}(A)$ is non-empty, every regular semisimple element $\gamma \in \mathbf{G}^{\mathrm{reg}}(k)$ has a quadratic splitting field $k(\gamma)$ over k. In other words, the centralizer $\mathbf{G}_\gamma$ is an anisotropic algebraic torus of rank one defined over k.

According to (1.6) and (1.7), we can identify

$$(5.1) \qquad \mathbf{G}(\mathbb{A}) = \mathrm{SL}_2(\mathbb{R})^a \times \mathrm{SL}_2(\mathbb{C})^b \times \mathrm{SU}_2(\mathbb{C})^c \times \prod_{\mathfrak{p}\in\mathrm{ram}_f(A)} \mathrm{SL}_1(D_\mathfrak{p}) \times \prod_{\mathfrak{p}\notin\mathrm{ram}_f(A)} \mathrm{SL}_2(k_\mathfrak{p}).$$

We specify the *standard measure* μ on $\mathbf{G}(\mathbb{A})$ (resp. $\mathbf{G}_\gamma(\mathbb{A})$). Let v be a place of k. For $v \in \mathrm{ram}(A)$, let μ_v be the Haar probability measure on $\mathbf{G}(k_v)$. For $v \notin \mathrm{ram}(A)$ archimedean (resp. non-archimedean), let μ_v be the Haar measure on $\mathbf{G}(k_v)$ defined in §3.1 (resp. §4.1). We define the global measure $\mu = \mu_\infty \times \mu_f$ on $\mathbf{G}(\mathbb{A}) = \mathbf{G}(\mathbb{A}_\infty) \times \mathbf{G}(\mathbb{A}_f)$ as the product of the local measures μ_v. By an abuse of notation, we also denote by μ_v the Haar measure on $\mathbf{G}_\gamma(k_v)$ defined in §3.2 and §4.1, and we define $\mu := \prod_v \mu_v$ on $\mathbf{G}_\gamma(\mathbb{A})$. Finally, we endow the lattices $\mathbf{G}(k)$ and $\mathbf{G}_\gamma(k)$ with counting measures and the corresponding quotients $\mathbf{G}(k)\backslash\mathbf{G}(\mathbb{A})$ and $\mathbf{G}_\gamma(k)\backslash\mathbf{G}_\gamma(\mathbb{A})$ with the unique compatible right $\mathbf{G}(\mathbb{A})$-invariant (resp. right $\mathbf{G}_\gamma(\mathbb{A})$-invariant) measures.

5.1. Borel's volume formula.

In this section, we record the classical covolume of the group of norm 1 elements of a maximal order of A. The result is really due to Borel [5, §7.3], but we differ from his formula by a factor of 2 as we work with special linear groups rather than projective ones. For clarity, we provide a detailed proof based on the local calculations in [5, pp. 17–18] and the fact that the *Tamagawa number* of $\mathbf{G}$ equals 1 [57, Th. 3.3.1]. As a supplement, we also provide an explicit lower bound for the covolume, based on Odlyzko's discriminant bound [45, Th. 1].

Proposition 4. *Let* $\Gamma := \Gamma_\emptyset(\mathfrak{o})$ *in the notation of* (1.10). *Then*

$$\mathrm{vol}(\Gamma\backslash G) = \frac{\zeta_k(2)\Delta_k^{3/2}}{2^{3b+2c}\pi^{a+2b+2c}} \prod_{\mathfrak{p}\in\mathrm{ram}_f(A)} (N(\mathfrak{p}) - 1).$$

Moreover, $\mathrm{vol}(\Gamma\backslash G) > e^{-7}$.

Proof. Let us consider, in the notation of (1.8)–(1.9),

$$V := \mathrm{SU}_2(\mathbb{C})^c \times K_\emptyset(\mathfrak{o}).$$

Then $\mathbf{G}(\mathbb{A}) = \mathbf{G}(k)(G \times V)$ by the strong approximation property [41, Th. 7.7.5], and hence

$$\mathbf{G}(k)\backslash\mathbf{G}(\mathbb{A})/V \simeq \eta(\mathbf{G}(k)\cap(G\times V))\backslash G = \Gamma\backslash G,$$

where $\eta\colon \mathbf{G}(\mathbb{A}) \to G$ is the projection introduced below (1.6). As V is a product of maximal compact subgroups over the relevant places of k, the definition of the standard measure yields

$$(5.2) \qquad\qquad \mu(\mathbf{G}(k)\backslash\mathbf{G}(\mathbb{A})) = \mathrm{vol}(\Gamma\backslash G).$$

It remains to calculate the left-hand side. The standard measure μ on $\mathbf{G}(\mathbb{A})$ is proportional to the *Tamagawa measure* $\tau = \Delta_k^{-3/2}\omega$, as defined in [5, §6.2]. Combining the two Lemmata on [5, pp. 17–18] with (3) on [5, p. 18], we see that $\mu = C\,\omega$, where

$$C^{-1} = \pi^a\left(8\pi^2\right)^b\left(4\pi^2\right)^c \prod_{\mathfrak{p}\in\mathrm{ram}_f(A)} \frac{N(\mathfrak{p})+1}{N(\mathfrak{p})^2} \prod_{\mathfrak{p}\notin\mathrm{ram}_f(A)} \frac{N(\mathfrak{p})^2-1}{N(\mathfrak{p})^2}.$$

On the other hand, $\tau(\mathbf{G}(k)\backslash\mathbf{G}(\mathbb{A})) = 1$ by [57, Th. 3.3.1], therefore

$$\mu(\mathbf{G}(k)\backslash\mathbf{G}(\mathbb{A})) = C\,\Delta_k^{3/2} = \frac{\zeta_k(2)\Delta_k^{3/2}}{2^{3b+2c}\pi^{a+2b+2c}} \prod_{\mathfrak{p}\in\mathrm{ram}_f(A)} (N(\mathfrak{p}) - 1).$$

Finally, we show that the right-hand side exceeds e^{-7}. For this we apply [45, Th. 1] with the parameters $\sigma := 3.399$ and $\tilde\sigma := 2.762$. The conditions [45, (1.5a)–(1.5b)] are satisfied, hence [45, (1.6)] yields the explicit bound

$$\Delta_k > 11.5948^{a+c}\ 7.6189^{2b}\ e^{-4.0497}.$$

We obtained the numerical values with SageMath 8.3, and verified them independently with Mathematica 10.1. Therefore,

$$\frac{\Delta_k^{3/2}}{2^{3b+2c}\pi^{a+2b+2c}} > \left(\frac{39.48}{4\pi^2}\right)^{a+c}\left(\frac{442.25}{8\pi^2}\right)^b e^{-6.0746}.$$

The fractions on the right-hand side exceed 1, hence we are done. $\qquad\square$

5.2. Volumes of centralizers.

In this section, we estimate $\mu(\mathbf{G}_\gamma(k)\backslash\mathbf{G}_\gamma(\mathbb{A}))$ for $\gamma \in \mathbf{G}^{\mathrm{reg}}(k)$. Before stating and proving the actual result, we collect some relevant facts about algebraic tori.

Let $\mathbf{G}_m$ be the multiplicative group, regarded as a torus of rank one defined over $\mathbb{Q}$. Let k be a number field, and let $\mathbf{T}$ be a torus of rank r defined over k. The character group

$$X^*(\mathbf{T}) := \mathrm{Hom}(\mathbf{T}_{\overline{\mathbb{Q}}}, \mathbf{G}_{m\overline{\mathbb{Q}}})$$

is a free abelian group of rank r. A Galois automorphism $\sigma \in \mathrm{Gal}(\mathbb{Q})$ determines the torus $\sigma\mathbf{T}$ of rank r defined over σk, and an isomorphism of abelian groups

$$\sigma\colon X^*(\mathbf{T}) \xrightarrow{\sim} X^*(\sigma\mathbf{T}).$$

The torus $\sigma\mathbf{T}$ only depends on the coset $\sigma\operatorname{Gal}(k)$, hence in particular $X^*(\mathbf{T}) \simeq \mathbb{Z}^r$ is a $\operatorname{Gal}(k)$-module. Similarly, $\mathbf{T}' := \operatorname{Res}_{k/\mathbb{Q}}\mathbf{T}$ is a torus of rank $r[k:\mathbb{Q}]$ defined over $\mathbb{Q}$, hence $X^*(\mathbf{T}') \simeq \mathbb{Z}^{r[k:\mathbb{Q}]}$ is a $\operatorname{Gal}(\mathbb{Q})$-module. In fact [43, §2.61] shows that

$$\mathbf{T}'_{\overline{\mathbb{Q}}} \simeq \prod_{\sigma \in \operatorname{Gal}(\mathbb{Q})/\operatorname{Gal}(k)} (\sigma\mathbf{T})_{\overline{\mathbb{Q}}},$$

therefore

$$X^*(\mathbf{T}') \simeq \prod_{\sigma \in \operatorname{Gal}(\mathbb{Q})/\operatorname{Gal}(k)} \sigma X^*(\mathbf{T}) \simeq \operatorname{Ind}_{\operatorname{Gal}(k)}^{\operatorname{Gal}(\mathbb{Q})} X^*(\mathbf{T}).$$

For an arbitrary place w of $\mathbb{Q}$, there is an isomorphism of k-algebras

$$\mathbb{Q}_w \otimes_{\mathbb{Q}} k \simeq \prod_{v|w} k_v,$$

where v runs through the places of k lying above w. Correspondingly, there is an isomorphism of topological groups

$$\mathbf{T}'(\mathbb{Q}_w) \simeq \prod_{v|w} \mathbf{T}(k_v).$$

Let us identify $\operatorname{Gal}(\mathbb{Q}_w)$ with a subgroup of $\operatorname{Gal}(\mathbb{Q})$ in the usual way. In the left action of $\operatorname{Gal}(\mathbb{Q})$ on $\operatorname{Gal}(\mathbb{Q})/\operatorname{Gal}(k)$, each orbit of $\operatorname{Gal}(\mathbb{Q}_w)$ corresponds to a place v of k lying above w, and the orbit itself can be identified with $\operatorname{Gal}(\mathbb{Q}_w)/\operatorname{Gal}(k_v)$. It follows that

$$X^*(\mathbf{T}')^{\operatorname{Gal}(\mathbb{Q}_w)} \simeq \prod_{v|w} X^*(\mathbf{T})^{\operatorname{Gal}(k_v)}.$$

This isomorphism commutes with evaluation of local characters, so the left-hand side as a subgroup of $\operatorname{Hom}(\mathbf{T}'(\mathbb{Q}_w), \mathbb{Q}_w^\times)$ corresponds to the right-hand side as a direct product of subgroups of $\operatorname{Hom}(\mathbf{T}(k_v), k_v^\times)$. This will show that the *standard measure* on $\mathbf{T}'(\mathbb{Q}_w)$, to be defined in the next paragraph, equals the product of the analogous standard measures on $\mathbf{T}(k_v)$.

Following [46, §2.1], the topological group $\mathbf{T}'(\mathbb{Q}_w)$ has the unique maximal compact subgroup

$$\mathbf{T}'(\mathbb{Q}_w)^\flat := \{t \in \mathbf{T}'(\mathbb{Q}_w) : |\xi(t)|_w = 1 \text{ for all } \xi \in X^*(\mathbf{T}')^{\operatorname{Gal}(\mathbb{Q}_w)}\}.$$

By choosing a(ny) basis $\{\xi_1, \ldots, \xi_s\}$ of $X^*(\mathbf{T}')^{\operatorname{Gal}(\mathbb{Q}_w)}$, we see that $\mathbf{T}'(\mathbb{Q}_w)/\mathbf{T}'(\mathbb{Q}_w)^\flat$ is isomorphic to $\mathbb{Z}^s$ when w is non-archimedean (resp. $\mathbb{R}^s$ when w is archimedean). Accordingly, we define the standard measure μ_w on $\mathbf{T}'(\mathbb{Q}_w)$ as the product of the Haar probability measure $\mu_w^\flat$ on $\mathbf{T}'(\mathbb{Q}_w)^\flat$ and the counting measure on $\mathbb{Z}^s$ (resp. Lebesgue measure on $\mathbb{R}^s$). Explicitly, when w is non-archimedean, μ_w is the unique Haar measure on $\mathbf{T}'(\mathbb{Q}_w)$ such that $\mu_w(\mathbf{T}'(\mathbb{Q}_w)^\flat) = 1$; when w is archimedean,

$$\int_{\mathbf{T}'(\mathbb{Q}_w)} f(t) \, d\mu_w(t) = \int_{\mathbf{T}'(\mathbb{Q}_w)/\mathbf{T}'(\mathbb{Q}_w)^\flat} \left(\int_{\mathbf{T}'(\mathbb{Q}_w)^\flat} f(t't'') \, d\mu_w^\flat(t'') \right) \frac{d|\xi_1(t')|_w}{|\xi_1(t')|_w} \cdots \frac{d|\xi_s(t')|_w}{|\xi_s(t')|_w}.$$

The *standard measure* μ on $\mathbf{T}'(\mathbb{A}_{\mathbb{Q}})$ is the product of the local measures μ_w. The standard measure μ on $\mathbf{T}(\mathbb{A}_k)$ is defined analogously, using the characters of $\mathbf{T}$ defined over the various completions k_v. By our earlier remarks, these two measures correspond to each other under the natural isomorphism $\mathbf{T}'(\mathbb{A}_{\mathbb{Q}}) \simeq \mathbf{T}(\mathbb{A}_k)$.

Proposition 5. *Let $\gamma \in \mathbf{G}^{\operatorname{reg}}(k)$ be a regular semisimple element. Then*

$$\mu(\mathbf{G}_\gamma(k) \backslash \mathbf{G}_\gamma(\mathbb{A})) \ll_\varepsilon \Delta_k^{1/2+\varepsilon} |N_{k/\mathbb{Q}}(\Delta_{k(\gamma)/k})|^{1/2+\varepsilon}.$$

Proof. Let us abbreviate $\mathbf{T} := \mathbf{G}_\gamma$ and $l := k(\gamma)$. By our initial remarks,

(5.3) $$\mu(\mathbf{T}(k) \backslash \mathbf{T}(\mathbb{A}_k)) = \mu(\mathbf{T}'(\mathbb{Q}) \backslash \mathbf{T}'(\mathbb{A}_{\mathbb{Q}})),$$

where $\mathbf{T}' := \operatorname{Res}_{k/\mathbb{Q}}\mathbf{T}$ is an anisotropic torus of rank $[k:\mathbb{Q}]$ defined over $\mathbb{Q}$. The standard measure μ on $\mathbf{T}'(\mathbb{A})$ is proportional to the *Tamagawa measure* ω on $\mathbf{T}'(\mathbb{A}_{\mathbb{Q}})$, as defined by [54, (12)]. The square of the proportionality constant is called the *quasi-discriminant* $D_{\mathbf{T}'}$ [54, (13)], that is, $\mu = D_{\mathbf{T}'}^{1/2}\omega$. Moreover, by the definition of the *Tamagawa number* $\tau_{\mathbf{T}'}$ [46, p. 126],

$$\omega(\mathbf{T}'(\mathbb{Q}) \backslash \mathbf{T}'(\mathbb{A}_{\mathbb{Q}})) = \tau_{\mathbf{T}'} \, \rho_{\mathbf{T}'},$$

where $\rho_{\mathbf{T}'}$ is the special value at $s = 1$ of the Artin L-function

$$L(s, \mathbf{T}') := L(s, X^*(\mathbf{T}') \otimes \mathbb{C}).$$

It follows that

$$\mu(\mathbf{T}'(\mathbb{Q}) \backslash \mathbf{T}'(\mathbb{A}_{\mathbb{Q}})) = \tau_{\mathbf{T}'} \, \rho_{\mathbf{T}'} \, D_{\mathbf{T}'}^{1/2}.$$

By [54, Prop. 3.1], the quasi-discriminant $D_{\mathbf{T}'}$ is smaller than the conductor $a_{\mathbf{T}'}$ of $L(s, \mathbf{T}')$. On the other hand, by our initial remarks, $L(s, \mathbf{T}')$ is the same as $L(s, \mathbf{T})$, hence $a_{\mathbf{T}'} = a_{\mathbf{T}}$ and $\rho_{\mathbf{T}'} = \rho_{\mathbf{T}}$. In addition, $\tau_{\mathbf{T}'} = \tau_{\mathbf{T}}$ by [46, Th. 3.5.1]. Recalling also (5.3), we conclude that

$$\mu(\mathbf{T}(k) \backslash \mathbf{T}(\mathbb{A}_k)) < \tau_{\mathbf{T}} \, \rho_{\mathbf{T}} \, a_{\mathbf{T}}^{1/2}.$$

The absolute Galois group $\mathrm{Gal}(k)$ acts on $X^*(\mathbf{T}) \simeq \mathbb{Z}$ by the nontrivial quadratic character χ fixing $\mathrm{Gal}(l)$. Therefore, by the Corollary on [47, p. 69] and an explicit calculation with cocycles and coboundaries (cf. [10, p. 97]),

$$\tau_{\mathbf{T}} = \#H^1(\mathrm{Gal}(l/k), X^*(\mathbf{T})) = 2.$$

Moreover,

$$L(s, \mathbf{T}) = L(s, \chi) = \zeta_l(s)/\zeta_k(s),$$

whose conductor equals

$$a_{\mathbf{T}} = \Delta_l/\Delta_k = \Delta_k \, |N_{k/\mathbb{Q}}(\Delta_{l/k})|.$$

Finally, by Lemma 4 in [40, Ch. XVI, §3],

$$\rho_{\mathbf{T}} = L(1, \chi) \ll_\varepsilon a_{\mathbf{T}}^\varepsilon.$$

Putting everything together,

$$\mu(\mathbf{T}(k) \backslash \mathbf{T}(\mathbb{A}_k)) \ll_\varepsilon a_{\mathbf{T}}^{1/2+\varepsilon} = \Delta_k^{1/2+\varepsilon} |N_{k/\mathbb{Q}}(\Delta_{l/k})|^{1/2+\varepsilon},$$

and we are done. $\qquad\square$

5.3. **Counting conjugacy classes.** Let us introduce the notation

$$k_A^+ := \{x \in k^\times : x_v > 0 \text{ for all } v \in \mathrm{ram}_\infty(A)\}.$$

Lemma 14. *Let $\gamma \in \mathbf{G}^{\mathrm{reg}}(k)$ be a regular semisimple element. There is a bijection*

$$\{[h^{-1}\gamma h] \colon h \in \mathbf{H}(k)\} \xrightarrow{\sim} k_A^+/N_{k(\gamma)/k}(k(\gamma)^\times).$$

Explicitly, the $\mathbf{G}(k)$-conjugacy class $[h^{-1}\gamma h]$ is mapped to the image of $\mathrm{n}(h)$ under the natural quotient map.

Proof. For arbitrary $h_1, h_2 \in \mathbf{H}(k)$, we can see that

$$
\begin{aligned}
[h_1^{-1}\gamma h_1] = [h_2^{-1}\gamma h_2] \qquad &\Longleftrightarrow \qquad \exists g \in \mathbf{G}(k) : h_1^{-1}\gamma h_1 = g^{-1}h_2^{-1}\gamma h_2 g \\
&\Longleftrightarrow \qquad \exists g \in \mathbf{G}(k) : h_1 h_2^{-1} \in \mathbf{H}_\gamma(k) h_2 g h_2^{-1} \\
&\Longleftrightarrow \qquad h_1 h_2^{-1} \in \mathbf{H}_\gamma(k) \mathbf{G}(k).
\end{aligned}
$$

So we can identify

$$\{[h^{-1}\gamma h] : h \in \mathbf{H}(k)\} \simeq \mathbf{H}(k)/(\mathbf{H}_\gamma(k)\mathbf{G}(k)) \simeq \mathrm{n}(\mathbf{H}(k))/\mathrm{n}(\mathbf{H}_\gamma(k)),$$

where the second bijection is a group isomorphism induced by the reduced norm. Let us abbreviate $l := k(\gamma)$ and write σ for the nontrivial Galois automorphism of l over k. Then $\mathrm{n}(\mathbf{H}(k)) = k_A^+$ by the Hasse–Schilling theorem [56, Ch. XI, Prop. 3], while $\mathrm{n}(\mathbf{H}_\gamma(k)) = N_{l/k}(l^\times)$, because $\mathbf{H}_\gamma(k)$ is conjugate to $\{\mathrm{diag}(\beta, \beta^\sigma) : \beta \in l^\times\}$ inside $\mathbf{H}(l) \simeq \mathrm{GL}_2(l)$. The result follows. $\qquad\square$

We recall the notation (1.8)–(1.9) and the decomposition (5.1). Let us also introduce

$$\mathcal{K} := \bigcup_{g \in \mathbf{G}(\mathbb{A}_f)} g K_\emptyset(\mathfrak{o}) g^{-1}.$$

Using Lemma 14, we can bound effectively the number of $\mathbf{G}(k)$-conjugacy classes which intersect $\mathcal{K}$ and lie inside a given $\mathbf{H}(k)$-conjugacy class.

Lemma 15. *Let $\gamma \in \mathbf{G}^{\mathrm{reg}}(k) \cap \mathcal{K}$ be a regular semisimple element. The cardinality of*

$$[\gamma]' := \{[h^{-1}\gamma h] : h \in \mathbf{H}(k) \ and \ h^{-1}\gamma h \in \mathcal{K}\}$$

is at most $2^{a+|\mathrm{ram}_f(A)|+\omega(\Delta(\gamma))}$, where $\omega(\Delta(\gamma))$ is the number of distinct prime ideals dividing $\Delta(\gamma)$.

Proof. Let us abbreviate $l := k(\gamma)$. By Lemma 14, we can think of $[\gamma]'$ as a subset of $k_A^+/N_{l/k}(l^\times)$. On the other hand, by the Hasse norm theorem [44, Ch. VI, Cor. 4.5], there is a natural embedding of groups

$$k_A^+/N_{l/k}(l^\times) \hookrightarrow \bigoplus_{v \notin \mathrm{ram}_\infty(A)} k_v^\times/N_{l_v/k_v}(l_v^\times),$$

where $l_v := l \otimes_k k_v$. The index $[k_v^\times : N_{l_v/k_v}(l_v^\times)]$ equals 2 or 1 depending on whether l_v is a field or not. Hence it suffices to verify that $\mathrm{n}(h) \in N_{l_{\mathfrak{p}}/k_{\mathfrak{p}}}(l_{\mathfrak{p}}^\times)$ always holds when:

- $\mathfrak{p}$ is a fixed prime such that $\mathfrak{p} \notin \mathrm{ram}_f(A)$, $\mathfrak{p} \nmid \Delta(\gamma)$, and $l_{\mathfrak{p}}$ is a field;
- $h \in \mathrm{GL}_2(k_{\mathfrak{p}})$ and $g \in \mathrm{SL}_2(k_{\mathfrak{p}})$ are such that $g^{-1}h^{-1}\gamma h g \in \mathrm{SL}_2(\mathfrak{o}_{\mathfrak{p}})$.

By the initial assumptions, there exist $i \in \mathrm{SL}_2(k_{\mathfrak{p}})$ and $\delta \in \mathrm{SL}_2(\mathfrak{o}_{\mathfrak{p}})$ such that $\gamma = i^{-1}\delta i$. Of course $\Delta(\gamma) = \Delta(\delta)$. The conditions imply that, in the Bruhat–Tits tree $\mathcal{X}_{k_{\mathfrak{p}}}$, the vertices v_0 and $(ihg)v_0$ are fixed by the regular semisimple element $\delta \in \mathrm{SL}_2(\mathfrak{o}_{\mathfrak{p}})$. From Lemma 13 we infer that the quadratic extension $l_{\mathfrak{p}}/k_{\mathfrak{p}}$ is unramified (which also follows more directly from $\mathfrak{p} \nmid \Delta(\gamma)$), and there is *exactly one* vertex fixed by δ. We conclude that $v_0 = (ihg)v_0$, hence $v_{\mathfrak{p}}(\mathrm{n}(ihg)) \in 2\mathbb{Z}$, and so $\mathrm{n}(h) = \mathrm{n}(ihg) \in N_{l_{\mathfrak{p}}/k_{\mathfrak{p}}}(l_{\mathfrak{p}}^\times)$ by [56, Ch. VIII, Prop. 3]. The proof is complete. $\qquad\square$

The next two simple estimates are similar in nature, and they will be used in conjunction with Lemma 15.

Lemma 16. *Let $\Gamma := \Gamma_\emptyset(\mathfrak{o})$ in the notation of (1.10). Then*

$$2^{|\mathrm{ram}_f(A)|} \ll_\varepsilon \mathrm{vol}(\Gamma \backslash G)^\varepsilon.$$

Proof. Let $\varepsilon > 0$ be arbitrary. By Proposition 4,

$$\prod_{\mathfrak{p} \in \mathrm{ram}_f(A)} 2 \ll_\varepsilon \prod_{\substack{\mathfrak{p} \in \mathrm{ram}_f(A) \\ N(\mathfrak{p}) \geqslant 1+2^{1/\varepsilon}}} 2 \leqslant \prod_{\mathfrak{p} \in \mathrm{ram}_f(A)} (N(\mathfrak{p})-1)^\varepsilon \ll_\varepsilon \mathrm{vol}(\Gamma \backslash G)^\varepsilon.$$

The proof is complete. $\qquad\square$

Lemma 17. *Let $\mathfrak{m} \subset \mathfrak{o}$ be a nonzero ideal, and let $\omega(\mathfrak{m})$ be the number of its prime ideal divisors. Then*

$$2^{\omega(\mathfrak{m})} \ll_\varepsilon N(\mathfrak{m})^\varepsilon.$$

Proof. Let $\varepsilon > 0$ be arbitrary. Then

$$\prod_{\mathfrak{p}|\mathfrak{m}} 2 \ll_\varepsilon \prod_{\substack{\mathfrak{p}|\mathfrak{m} \\ N(\mathfrak{p}) \geqslant 2^{1/\varepsilon}}} 2 \leqslant \prod_{\mathfrak{p}|\mathfrak{m}} N(\mathfrak{p})^\varepsilon \leqslant N(\mathfrak{m})^\varepsilon.$$

The proof is complete. $\qquad\square$

6. Proof of Theorem 1

In this section, we prove Theorem 1. Recall the parameters $S \subset S_\infty^G$, $\boldsymbol{\sigma} = (\sigma_j) \in [0, 1/2]^S$, and $\boldsymbol{T} = (T_j) \in \mathbb{R}^{S_\infty^G \setminus S}$.

In §6.1, we construct a strongly positive definite test function $f_{\mathbb{A}} \in C_c(\mathbf{G}(\mathbb{A}))$. In §6.2, we show that the spectral side of the trace formula for the associated integral operator $Rf_{\mathbb{A}}$ detects with high weights the multiplicities $\mathrm{m}(\pi, \Gamma_\kappa(\mathfrak{n}))$ for $\pi \in \mathcal{B}(\boldsymbol{\sigma}, \boldsymbol{T})$. In §6.3, we estimate the contributions of individual conjugacy classes to the geometric side of the trace formula for $Rf_{\mathbb{A}}$. In §6.4, we combine the bounds from the spectral and geometric expansions with a crucial geometry of numbers ingredient from [20] to conclude the proof of Theorem 1.

6.1. **Test function.** Our construction of the test function $f_{\mathbb{A}} \in C_c(\mathbf{G}(\mathbb{A}))$ will depend on an additional family of nonnegative parameters $\boldsymbol{R} = (R_j) \in \mathbb{R}_{\geqslant 0}^S$, to be specified later in the proof of Theorem 1.

For $j \in S_\infty^G$, let (G_j, K_j) be the pair $(\mathrm{SL}_2(\mathbb{R}), \mathrm{SO}_2(\mathbb{R}))$ for $j \in \{1, \ldots, a\}$ and the pair $(\mathrm{SL}_2(\mathbb{C}), \mathrm{SU}_2(\mathbb{C}))$ for $j \in \{a+1, \ldots, a+b\}$. Then, $G = \prod_{j=1}^{a+b} G_j$ is the Lie group in Theorem 1, and $K := \prod_{j=1}^{a+b} K_j$ is a maximal compact subgroup as in §1.3.1. For $j \in S$, let $f_j \in C_c(K_j\backslash G_j/K_j)$ be the spherical function afforded by Proposition 1 for $R = R_j$. For $j \in S_\infty^G \setminus S$, let $f_j \in C_c(K_j\backslash G_j/K_j)$ be the spherical function afforded by Proposition 2 for $t = T_j$. We define

$$f := \bigotimes_{j=1}^{a+b} f_j \in C_c(K\backslash G/K),$$

as well as

(6.1) $$f_{\mathbb{A}} := f \otimes \mathbf{1}_V/\mu_f(K_\kappa(\mathfrak{n})), \qquad V := \mathrm{SU}_2(\mathbb{C})^c \times K_\kappa(\mathfrak{n}),$$

where μ_f is the standard measure on $\mathbf{G}(\mathbb{A}_f)$ from the beginning of §5.

By Propositions 1 and 2, the function $f_{\mathbb{A}} \in C_c(\mathbf{G}(\mathbb{A}))$ is supported in $\mathcal{D}(\boldsymbol{R}) \times V$, where

(6.2) $$\mathcal{D}(\boldsymbol{R}) := \prod_{j \in S} B(2R_j + 2) \times \prod_{j \in S_\infty^G \setminus S} B(2),$$

with $B(R)$ as in (3.19) and with the obvious ordering of the factors. Moreover,

(6.3) $$f_{\mathbb{A}}(g) \ll \mu_f(K_\kappa(\mathfrak{n}))^{-1} \prod_{j \in S_\infty^G \setminus S} (1 + |T_j|)^{\rho_j}, \qquad g \in \mathbf{G}(\mathbb{A}).$$

6.2. **Spectral side of the trace formula.** Let $Rf_{\mathbb{A}} \colon L^2(\mathbf{G}(k)\backslash \mathbf{G}(\mathbb{A})) \to L^2(\mathbf{G}(k)\backslash \mathbf{G}(\mathbb{A}))$ be the positive trace class operator given by

$$Rf_{\mathbb{A}}\phi(h) := \int_{\mathbf{G}(\mathbb{A})} \phi(hg) f_{\mathbb{A}}(g)\, dg.$$

The image of this operator is contained in the space $L^2(\mathbf{G}(k)\backslash \mathbf{G}(\mathbb{A}))^{K \times V}$ of vectors fixed by

$$K \times V = \mathrm{SO}_2(\mathbb{R})^a \times \mathrm{SU}_2(\mathbb{C})^b \times \mathrm{SU}_2(\mathbb{C})^c \times K_\kappa(\mathfrak{n}),$$

and so

(6.4) $$\operatorname{tr} Rf_{\mathbb{A}} = \operatorname{tr} Rf_{\mathbb{A}}|_{L^2(\mathbf{G}(k)\backslash \mathbf{G}(\mathbb{A}))^{K \times V}}.$$

We can identify $L^2(\mathbf{G}(k)\backslash \mathbf{G}(\mathbb{A}))^{K \times V}$ with a classical function space on $\Gamma_\kappa(\mathfrak{n})\backslash G/K$ as in (6.5) below. Indeed, $\mathbf{G}(\mathbb{A}) = \mathbf{G}(k)(G \times V)$ by the strong approximation property [41, Th. 7.7.5], and hence

$$\mathbf{G}(k)\backslash \mathbf{G}(\mathbb{A})/V \simeq \eta(\mathbf{G}(k) \cap (G \times V))\backslash G = \Gamma_\kappa(\mathfrak{n})\backslash G,$$

where $\eta \colon \mathbf{G}(\mathbb{A}) \to G$ is the projection introduced below (1.6). It follows that $L^2(\mathbf{G}(k)\backslash \mathbf{G}(\mathbb{A}))^V$ is isomorphic to $L^2(\Gamma_\kappa(\mathfrak{n})\backslash G)$, and

(6.5) $$L^2(\mathbf{G}(k)\backslash \mathbf{G}(\mathbb{A}))^{K \times V} \simeq L^2(\Gamma_\kappa(\mathfrak{n})\backslash G)^K \simeq \widehat{\bigoplus_{\boldsymbol{s}}} \mathrm{m}(\pi_{\boldsymbol{s}}, \Gamma_\kappa(\mathfrak{n}))\pi_{\boldsymbol{s}}^K,$$

recalling the parametrization of $\pi_{\boldsymbol{s}} \in \widehat{G}_{\mathrm{sph}}$ by $\boldsymbol{s} \in ((0, 1/2] \cup i[0, \infty))^{a+b}$ from §1.3.3. The action of $Rf_{\mathbb{A}}$ on the left-hand side becomes the action of the classical operator Rf (cf. (2.2)) on the space in the middle. From (3.16), it follows that this action becomes multiplication by the scalar

$$\widehat{f}(\boldsymbol{s}) := \prod_{j=1}^{a+b} \widehat{f}_j(s_j)$$

on each of the one-dimensional Hilbert space constituents $\pi_{\boldsymbol{s}}^K \simeq \bigotimes_{j=1}^{a+b} \pi_{s_j}^{K_j}$ on the right-hand side. Using also (6.4), we arrive at

(6.6) $$\operatorname{tr} Rf_{\mathbb{A}} = \sum_{\boldsymbol{s}} \mathrm{m}(\pi_{\boldsymbol{s}}, \Gamma_\kappa(\mathfrak{n}))\widehat{f}(\boldsymbol{s}),$$

where $\boldsymbol{s}$ runs through the same countable subset of $((0, 1/2] \cup i[0, \infty))^{a+b}$ as in (6.5). The right-hand side of (6.6) expresses the spectral side of the trace formula for $f_{\mathbb{A}}$.

For each $\pi_{\boldsymbol{s}}$ that occurs in (6.6), we have $\widehat{f}(\boldsymbol{s}) \geqslant 0$ by the positivity of $Rf_{\mathbb{A}}$. Moreover, Propositions 1 and 2 yield

$$\widehat{f}(\boldsymbol{s}) \gg \prod_{j \in S} e^{2\rho_j R_j \sigma_j}, \qquad \pi_{\boldsymbol{s}} \in \mathcal{B}(\boldsymbol{\sigma}, \boldsymbol{T}).$$

Therefore,

$$(6.7) \qquad \sum_{\pi \in \mathcal{B}(\boldsymbol{\sigma}, \boldsymbol{T})} \mathrm{m}(\pi, \Gamma_\kappa(\mathfrak{n})) \ll e^{-\sum_{j \in S} 2\rho_j R_j \sigma_j} \operatorname{tr} Rf_{\mathbb{A}}.$$

6.3. **Geometric side of the trace formula.** By the trace formula for compact quotients [2, Part I, §1], we also have the geometric expansion

$$(6.8) \qquad \operatorname{tr} Rf_{\mathbb{A}} = \sum_{[\gamma] \subset \mathbf{G}(k)} \mu(\mathbf{G}_\gamma(k) \backslash \mathbf{G}_\gamma(\mathbb{A})) \mathbf{O}(\gamma, f_{\mathbb{A}}),$$

where $[\gamma]$ runs through the conjugacy classes of $\mathbf{G}(k)$, μ is the standard measure on $\mathbf{G}_\gamma(\mathbb{A})$ defined in the beginning of §5, and

$$\mathbf{O}(\gamma, f_{\mathbb{A}}) := \int_{\mathbf{G}_\gamma(\mathbb{A}) \backslash \mathbf{G}(\mathbb{A})} f_{\mathbb{A}}(g^{-1} \gamma g) \, dg$$

is the global orbital integral. Separating the contribution of the central and regular semisimple conjugacy classes, we obtain

$$(6.9) \qquad \operatorname{tr} Rf_{\mathbb{A}} = \mu(\mathbf{G}(k) \backslash \mathbf{G}(\mathbb{A}))(f_{\mathbb{A}}(\mathrm{id}) + f_{\mathbb{A}}(-\mathrm{id})) + \sum_{[\gamma] \in W} \mu(\mathbf{G}_\gamma(k) \backslash \mathbf{G}_\gamma(\mathbb{A})) \mathbf{O}(\gamma, f_{\mathbb{A}}),$$

where $W := W(\boldsymbol{R}, K_\kappa(\mathfrak{n}))$ is the set of conjugacy classes $[\gamma] \subset \mathbf{G}^{\mathrm{reg}}(k)$ such that the conjugacy class of γ in $\mathbf{G}(\mathbb{A})$ intersects $\mathcal{D}(\boldsymbol{R}) \times V$ (cf. (6.1)–(6.2)). Combining (5.2) for $\Gamma = \Gamma_\emptyset(\mathfrak{o})$ and (6.3), it is clear that

$$\mu(\mathbf{G}(k) \backslash \mathbf{G}(\mathbb{A})) f_{\mathbb{A}}(\pm \mathrm{id}) \ll \operatorname{vol}(\Gamma_\emptyset(\mathfrak{o}) \backslash G) \mu_f(K_\kappa(\mathfrak{n}))^{-1} \prod_{j \in S_\infty^G \backslash S} (1 + |T_j|)^{\rho_j}.$$

Here $\mu_f(K_\kappa(\mathfrak{n}))^{-1}$ is the same as $[K_\emptyset(\mathfrak{o}) : K_\kappa(\mathfrak{n})] = [\Gamma_\emptyset(\mathfrak{o}) : \Gamma_\kappa(\mathfrak{n})]$, hence the right-hand side equals $\mathcal{C}(\Gamma_\kappa(\mathfrak{n}), \boldsymbol{T})$ by (1.14). Going back to (6.9), we infer that

$$(6.10) \qquad \operatorname{tr} Rf_{\mathbb{A}} \ll \mathcal{C}(\Gamma_\kappa(\mathfrak{n}), \boldsymbol{T}) + \left| \sum_{[\gamma] \in W} \mu(\mathbf{G}_\gamma(k) \backslash \mathbf{G}_\gamma(\mathbb{A})) \mathbf{O}(\gamma, f_{\mathbb{A}}) \right|.$$

Note that $\operatorname{tr}(\gamma) \in \mathfrak{o}$ holds for every $[\gamma] \in W$, because γ lies in a maximal compact subgroup of $\mathbf{G}(\mathbb{A}_f)$. In particular, for every prime ideal $\mathfrak{p} \notin \operatorname{ram}(A)$, γ is $\mathrm{SL}_2(k_\mathfrak{p})$-conjugate to an element of $\mathrm{SL}_2(\mathfrak{o}_p)$. This will be important in the proof of Lemma 18 below, when we apply Proposition 3 in conjunction with Lemma 7.

In order to estimate $\mathbf{O}(\gamma, f_{\mathbb{A}})$ for $[\gamma] \in W$, we write $f_{\mathbb{A}}$ in the fully factorized form

$$(6.11) \qquad f_{\mathbb{A}} = \bigotimes_v f_v,$$

where f_v is as in §6.1 for $v \notin \operatorname{ram}(A)$ archimedean, $f_v := \mathbf{1}_{\mathbf{G}(k_v)}$ for $v \in \operatorname{ram}(A)$ arbitrary, $f_\mathfrak{p} := \mathbf{1}_{K_{\kappa(\mathfrak{p})}(\mathfrak{n}_\mathfrak{p})} / \mu_\mathfrak{p}(K_{\kappa(\mathfrak{p})}(\mathfrak{n}_\mathfrak{p}))$ for $\mathfrak{p} \mid \mathfrak{n}$, and $f_\mathfrak{p} := \mathbf{1}_{\mathrm{SL}_2(\mathfrak{o}_\mathfrak{p})}$ for $\mathfrak{p} \notin \operatorname{ram}(A)$ satisfying $\mathfrak{p} \nmid \mathfrak{n}$. Then we have the decomposition

$$(6.12) \qquad \mathbf{O}(\gamma, f_{\mathbb{A}}) = \prod_v \mathbf{O}'(\gamma, f_v), \qquad \mathbf{O}'(\gamma, f_v) := |\Delta(\gamma)|_v^{1/2} \, \mathbf{O}(\gamma, f_v).$$

Indeed, the normalization factors $|\Delta(\gamma)|_v^{1/2}$ cancel out by Artin's product formula [56, Ch. IV, Th. 5].

Further, it will be convenient to bound $\mathbf{O}(\gamma, f_{\mathbb{A}})$ in terms of $\operatorname{tr}(\gamma)$, hence in accordance with (4.4) and (4.5), we introduce the functions $w_j^{\mathfrak{n}, \mathfrak{p}} \colon \mathfrak{o}_\mathfrak{p} \to \mathbb{R}_{\geqslant 0}$ as

$$(6.13) \qquad w_0^{\mathfrak{n}, \mathfrak{p}}(x) := \begin{cases} N(\mathfrak{p})^{\min(v_\mathfrak{p}(x^2 - 4)/2, \lfloor v_\mathfrak{p}(\mathfrak{n})/2 \rfloor)}, & x^2 - 4 = \square, \\ N(\mathfrak{p})^{\lfloor v_\mathfrak{p}(\mathfrak{n})/2 \rfloor} \mathbf{1}_{v_\mathfrak{p}(x^2 - 4) \geqslant v_\mathfrak{p}(\mathfrak{n})}, & x^2 - 4 \neq \square; \end{cases}$$

$$(6.14) \qquad w_1^{\mathfrak{n}, \mathfrak{p}}(x) := N(\mathfrak{p})^{2 v_\mathfrak{p}(\mathfrak{n})} \mathbf{1}_{v_\mathfrak{p}(x^2 - 4) \geqslant 2 v_\mathfrak{p}(\mathfrak{n})}.$$

Thus, for a regular semisimple $\gamma \in \mathrm{SL}_2(\mathfrak{o}_{\mathfrak{p}})$, we have $w_j^{\mathfrak{n},\mathfrak{p}}(\mathrm{tr}\,\gamma) = w_j^{v_{\mathfrak{p}}(\mathfrak{n})}(\gamma)$ in the notation of (4.4) and (4.5). Note that, when $\mathfrak{p} \nmid \mathfrak{n}$, $w_j^{\mathfrak{n},\mathfrak{p}}(x) = 1$ for all $x \in \mathfrak{o}$. We also define the function $w_\kappa^{\mathfrak{n}} \colon \widehat{\mathfrak{o}} \to \mathbb{R}_{\geqslant 0}$ by

$$(6.15) \qquad w_\kappa^{\mathfrak{n}}(x) := \prod_{\mathfrak{p} \mid \mathfrak{n}} w_{\kappa(\mathfrak{p})}^{\mathfrak{n},\mathfrak{p}}(x_{\mathfrak{p}}).$$

Our estimate on the invariant orbital integrals in (6.12) is as follows. We emphasize that the constant $C \geqslant 4$ in the statement is absolute, independent of all data.

Lemma 18. *There exists an absolute constant $C \geqslant 4$ such that the following holds. Let $f_{\mathbb{A}} = \bigotimes_v f_v$ with $f_v \in C_c(\mathbf{G}(k_v))$ be as in (6.11), and assume that $[\gamma] \in W$. Then:*

(a) For every archimedean place v of k, $|\mathbf{O}'(\gamma, f_v)| \leqslant C$.

(b) For every non-archimedean place $\mathfrak{p}$ of k,

$$\mathbf{O}'(\gamma, f_{\mathfrak{p}}) \leqslant C |2|_{\mathfrak{p}}^{-1} |\Delta_{k(\gamma)/k}|_{\mathfrak{p}}^{1/2} \cdot \begin{cases} w_{\kappa(\mathfrak{p})}^{\mathfrak{n},\mathfrak{p}}(\mathrm{tr}\,\gamma), & \mathfrak{p} \mid \mathfrak{n}; \\ 1, & \mathfrak{p} \nmid \mathfrak{n}. \end{cases}$$

Moreover, if $\mathfrak{p} \nmid \Delta(\gamma)$, then the leading factor $C|2|_{\mathfrak{p}}^{-1}|\Delta_{k(\gamma)/k}|_{\mathfrak{p}}^{1/2}$ can be omitted.

(c) The global orbital integral satisfies

$$|\mathbf{O}(\gamma, f_{\mathbb{A}})| \leqslant C^{[k:\mathbb{Q}]+\omega(\Delta(\gamma))} |N_{k/\mathbb{Q}}(\Delta_{k(\gamma)/k})|^{-1/2} w_\kappa^{\mathfrak{n}}(\mathrm{tr}\,\gamma),$$

where $\omega(\Delta(\gamma))$ is the number of distinct prime ideals dividing $\Delta(\gamma)$.

Proof. Let v be an archimedean place of k. If $v \notin \mathrm{ram}(A)$, then the bound in (a) follows from Propositions 1 and 2, where the implied constants are absolute. If $v \in \mathrm{ram}(A)$, then $\mathbf{O}'(\gamma, f_v) = |\Delta(\gamma)|_v^{1/2} < 2$, since γ lies in $\mathbf{G}(k_v) \simeq \mathrm{SU}_2(\mathbb{C})$. This proves (a).

Before turning to the proof of (b), we make two initial observations.

- First, $|\Delta_{k(\gamma)/k}|_{\mathfrak{p}}$ equals 1 or $N(\mathfrak{p})^{-1}|4|_{\mathfrak{p}}$ depending on whether $\mathfrak{p}$ is unramified or ramified in $k(\gamma)/k$; this follows from [56, Ch. VIII, Prop. 3], [56, Ch. I, Prop. 4], and the Corollary to [56, Ch. VIII, Prop. 6].

- Second, $\Delta_{k(\gamma)/k} \mid \Delta(\gamma)$, and hence $|\Delta(\gamma)|_{\mathfrak{p}} \leqslant |\Delta_{k(\gamma)/k}|_{\mathfrak{p}}$ for every $\mathfrak{p}$, because $\Delta(\gamma)$ is the discriminant of the $\mathfrak{o}$-submodule $\mathfrak{o} + \mathfrak{o}w \subset \mathfrak{o}_{k(\gamma)}$, where w is an eigenvalue of γ.

Now, if $\mathfrak{p} \notin \mathrm{ram}(A)$, then the bounds in (b) follow from Lemma 7, Proposition 3, and our first observation. We note that the factor $\mathrm{vol}(K_j(\mathfrak{p}^r))$ in Proposition 3 is neutralized by the factor $1/\mu_{\mathfrak{p}}(K_{\kappa(\mathfrak{p})}(\mathfrak{n}_{\mathfrak{p}}))$ included in the definition of $f_{\mathfrak{p}}$ for $\mathfrak{p} \mid \mathfrak{n}$ (see below (6.11)). On the other hand, if $\mathfrak{p} \in \mathrm{ram}(A)$, then $\mathfrak{p} \nmid \mathfrak{n}$, and $\mathbf{O}'(\gamma, f_{\mathfrak{p}}) = |\Delta(\gamma)|_{\mathfrak{p}}^{1/2} \leqslant |\Delta_{k(\gamma)/k}|_{\mathfrak{p}}^{1/2}$ follows from our second observation and the fact that $\mathbf{G}_\gamma(k_{\mathfrak{p}}) \backslash \mathbf{G}(k_{\mathfrak{p}})$ is a probability space.

Part (c) follows directly from (6.12) and the already established parts (a)–(b), with $2C$ as the absolute constant. $\square$

Lemma 18 combined with Proposition 5 yields the following estimate on the contribution of each regular conjugacy class to the geometric side of the trace formula:

$$(6.16) \qquad \mu(\mathbf{G}_\gamma(k) \backslash \mathbf{G}_\gamma(\mathbb{A})) \mathbf{O}(\gamma, f_{\mathbb{A}}) \ll_\varepsilon C^{\omega(\Delta(\gamma))} \Delta_k^{1/2+\varepsilon} |N_{k/\mathbb{Q}}(\Delta_{k(\gamma)/k})|^\varepsilon w_\kappa^{\mathfrak{n}}(\mathrm{tr}\,\gamma), \qquad [\gamma] \in W.$$

From now on we assume that $\boldsymbol{R} = (R_j)_{j \in S}$ satisfies

$$(6.17) \qquad \sum_{j \in S} \rho_j R_j \leqslant 7 + \log \mathcal{C}(\Gamma_\kappa(\mathfrak{n}), \boldsymbol{T}),$$

where the right-hand side is positive by Proposition 4. In fact, with our final choice (6.31), we will have equality in (6.17) except when $S = \emptyset$, in which case (6.17) is vacuous. We may use (6.17) to consolidate the ε-powers in (6.16) as follows. The condition $[\gamma] \in W$ implies readily via Lemma 5 that

$$(6.18) \qquad \begin{aligned} &|\mathrm{tr}(\gamma)|_j < e^{\rho_j(R_j+2)}, & &|\Delta(\gamma)|_j < e^{2\rho_j(R_j+2)}, & &j \in S; \\ &|\mathrm{tr}(\gamma)|_j < e^{2\rho_j}, & &|\Delta(\gamma)|_j < e^{4\rho_j}, & &j \in \{1, \ldots, a+b+c\} \setminus S. \end{aligned}$$

The bounds (6.18) combined with (6.17) and the observation from the proof of Lemma 18 that $\Delta_{k(\gamma)/k} \mid \Delta(\gamma)$ show that

$$|N_{k/\mathbb{Q}}(\Delta_{k(\gamma)/k})| \leqslant |N_{k/\mathbb{Q}}(\Delta(\gamma))| \ll \mathcal{C}(\Gamma_\kappa(\mathfrak{n}), \boldsymbol{T})^2. \tag{6.19}$$

Moreover, by Proposition 4,

$$\Delta_k \ll \mathcal{C}(\Gamma_\kappa(\mathfrak{n}), \boldsymbol{T})^{2/3}. \tag{6.20}$$

Combining (6.10), (6.16), (6.19), (6.20), and Lemma 17, we conclude that

$$\operatorname{tr} Rf_{\mathbb{A}} \preccurlyeq \mathcal{C}(\Gamma_\kappa(\mathfrak{n}), \boldsymbol{T}) + \Delta_k^{1/2} \sum_{[\gamma] \in W} w_\kappa^{\mathfrak{n}}(\operatorname{tr} \gamma). \tag{6.21}$$

As before, $X \preccurlyeq Y$ stands for $X \ll_\varepsilon \mathcal{C}(\Gamma, \boldsymbol{T})^\varepsilon Y$. Our estimate on the geometric side of the trace formula (6.21) should be compared with (6.7) and (6.17).

6.4. **Geometry of numbers.** We may further estimate the right-hand side of (6.21) as follows. We have seen in (6.18) that the traces of $[\gamma] \in W$ lie in $\mathfrak{o} \cap \mathcal{R}$, where

$$\mathcal{R} := \left\{ x \in \mathbb{A}_\infty : |x|_j < e^{\rho_j(R_j+2)} \text{ for all } j \in S \text{ and } |x|_j < e^{2\rho_j} \text{ for all } j \in \{1, \ldots, a+b+c\} \setminus S \right\}. \tag{6.22}$$

For a given conjugacy class $[\gamma] \in W$, there are at most $2^{a+|\mathrm{ram}_f(A)|+\omega(\Delta(\gamma))}$ conjugacy classes in $\mathbf{G}(k)$ with the same trace $\operatorname{tr}(\gamma)$, as follows from the Skolem–Noether theorem [41, Cor. 2.9.9] and Lemma 15. Moreover, $2^{|\mathrm{ram}_f(A)|+\omega(\Delta(\gamma))} \preccurlyeq 1$ by Lemmata 16–17 and (6.19). Therefore, by counting how many times a trace appears in (6.21), we arrive at

$$\operatorname{tr} Rf_{\mathbb{A}} \preccurlyeq \mathcal{C}(\Gamma_\kappa(\mathfrak{n}), \boldsymbol{T}) + \Delta_k^{1/2} \sum_{x \in \mathfrak{o} \cap \mathcal{R}} w_\kappa^{\mathfrak{n}}(x). \tag{6.23}$$

We claim that the function $w_\kappa^{\mathfrak{n}}$ is periodic by the ideal (embedded into $\widehat{\mathfrak{o}}$)

$$\mathfrak{n}' := 4 \prod_{\mathfrak{p} \mid \mathfrak{n}} \mathfrak{p}^{(1+\kappa(\mathfrak{p}))v_{\mathfrak{p}}(\mathfrak{n})}. \tag{6.24}$$

By (6.15), it suffices to show for every $\mathfrak{p} \mid \mathfrak{n}$ that $w_0^{\mathfrak{n},\mathfrak{p}}$ is periodic modulo $4\mathfrak{p}^{v_{\mathfrak{p}}(\mathfrak{n}_{\mathfrak{p}})}$ and $w_1^{\mathfrak{n},\mathfrak{p}}$ is periodic modulo $\mathfrak{p}^{2v_{\mathfrak{p}}(\mathfrak{n}_{\mathfrak{p}})}$. For this, we first rewrite the definition (6.13) in the more transparent form

$$w_0^{\mathfrak{n},\mathfrak{p}}(x) = N(\mathfrak{p})^{v_{\mathfrak{p}}(x^2-4)/2} \mathbf{1}_{v_{\mathfrak{p}}(x^2-4) < v_{\mathfrak{p}}(\mathfrak{n})} \mathbf{1}_{x^2-4=\square} + N(\mathfrak{p})^{\lfloor v_{\mathfrak{p}}(\mathfrak{n})/2 \rfloor} \mathbf{1}_{v_{\mathfrak{p}}(x^2-4) \geqslant v_{\mathfrak{p}}(\mathfrak{n})}. \tag{6.25}$$

Using also the definition (6.14), it suffices to check the following three claims for any given $r \in \mathbb{N}$:

- for $x \in \mathfrak{o}_{\mathfrak{p}}$, the condition $v_{\mathfrak{p}}(x^2-4) \geqslant r$ is invariant under shifts by elements of $\mathfrak{p}^r$;
- for $x \in \mathfrak{o}_{\mathfrak{p}}$ satisfying $v_{\mathfrak{p}}(x^2-4) < r$, the quantity $v_{\mathfrak{p}}(x^2-4)$ is periodic by $\mathfrak{p}^r$;
- for $x \in \mathfrak{o}_{\mathfrak{p}}$ satisfying $v_{\mathfrak{p}}(x^2-4) < r$, the condition $x^2-4=\square$ is invariant under shifts by elements of $4\mathfrak{p}^r$.

The first two claims are clear, so we focus on the third one. For $\mathfrak{p} \nmid 2$, the required invariance follows from the fact that the square map is an automorphism of $1 + \mathfrak{p}\mathfrak{o}_{\mathfrak{p}}$ (see [56, Ch. II, Prop. 8]). For $\mathfrak{p} \mid 2$, the required invariance follows from the fact that the square map is an isomorphism from $1 + 4\mathfrak{o}_{\mathfrak{p}}$ to $1 + 8\mathfrak{o}_{\mathfrak{p}}$ (see the proof of [56, Ch. II, Prop. 9]).

Now we can estimate the right-hand side of (6.23) with the help of the following lemma.

Lemma 19. *Let $w \colon \widehat{\mathfrak{o}} \to \mathbb{R}_{\geqslant 0}$ be a continuous function periodic by a nonzero ideal $\mathfrak{m} \subset \mathfrak{o}$, and let $\mathcal{P} \subset \mathbb{A}_\infty$ be a polycylinder of the form*

$$\mathcal{P} = \left\{ x \in \mathbb{A}_\infty : |x_v|_v \leqslant P_v \text{ for every archimedean place } v \text{ of } k \right\}.$$

Then,

$$\sum_{x \in \mathfrak{o} \cap \mathcal{P}} w(x) \ll \left(\Delta_k^{1/2} N(\mathfrak{m}) + \Delta_k^{-1/2} \operatorname{vol}(\mathcal{P}) \right) \int_{\widehat{\mathfrak{o}}} w.$$

Proof. By the periodicity condition, w is a linear combination (with unique nonnegative coefficients) of the characteristic functions of the cosets of $\mathfrak{m}\widehat{\mathfrak{o}}$ in $\widehat{\mathfrak{o}}$. Therefore, by linearity, it suffices to verify the inequality when w is one of these characteristic functions, in which case it states that

$$\#\big((y+\mathfrak{m})\cap\mathcal{P}\big)\ll\Delta_k^{1/2}+\frac{\operatorname{vol}(\mathcal{P})}{\Delta_k^{1/2}N(\mathfrak{m})},\qquad y\in\mathfrak{o}.$$

The claim is trivial when $(y+\mathfrak{m})\cap\mathcal{P}=\emptyset$. Otherwise, let us fix a point $x_0\in(y+\mathfrak{m})\cap\mathcal{P}$. For any $x\in(y+\mathfrak{m})\cap\mathcal{P}$, we have $x-x_0\in\mathfrak{m}\cap(2\mathcal{P})$, hence it suffices to show that

$$\#\big(\mathfrak{m}\cap(2\mathcal{P})\big)\ll\Delta_k^{1/2}+\frac{\operatorname{vol}(\mathcal{P})}{\Delta_k^{1/2}N(\mathfrak{m})}.$$

This inequality is equivalent to [20, Cor. 1], hence we are done. (Strictly speaking, [20] is written for $k\neq\mathbb{Q}$, but the conclusions extend trivially to $k=\mathbb{Q}$.) $\qquad\square$

Combining (6.24) with Proposition 4 and Lemma 10, we verify that

$$N(\mathfrak{n}')\ll\prod_{\mathfrak{p}|\mathfrak{n}}N(\mathfrak{p})^{(1+\kappa(\mathfrak{p}))v_\mathfrak{p}(\mathfrak{n})}\ll\Delta_k^{-3/2}\mathcal{C}(\Gamma_\kappa(\mathfrak{n}),\boldsymbol{T});\tag{6.26}$$

moreover, (6.22) and (6.17) yield $\operatorname{vol}(\mathcal{R})\ll\mathcal{C}(\Gamma_\kappa(\mathfrak{n}),\boldsymbol{T})$. Therefore, Lemma 19 furnishes the clean bound

$$\sum_{x\in\mathfrak{o}\cap\mathcal{R}}w_\kappa^\mathfrak{n}(x)\ll\Delta_k^{-1/2}\mathcal{C}(\Gamma_\kappa(\mathfrak{n}),\boldsymbol{T})\int_{\widehat{\mathfrak{o}}}w_\kappa^\mathfrak{n}.\tag{6.27}$$

The integral over $\widehat{\mathfrak{o}}$ splits into a product of local integrals as

$$\int_{\widehat{\mathfrak{o}}}w_\kappa^\mathfrak{n}=\prod_{\mathfrak{p}|\mathfrak{n}}\int_{\mathfrak{o}_\mathfrak{p}}w_{\kappa(\mathfrak{p})}^{\mathfrak{n},\mathfrak{p}},$$

and the factors can be estimated as (cf. (6.25) and (6.14))

$$\int_{\mathfrak{o}_\mathfrak{p}}w_0^{\mathfrak{n},\mathfrak{p}}\leqslant\int_{\mathfrak{o}_\mathfrak{p}}\left(\sum_{0\leqslant\ell<v_\mathfrak{p}(\mathfrak{n})/2}N(\mathfrak{p})^\ell\mathbf{1}_{v_\mathfrak{p}(x^2-4)=2\ell}+N(\mathfrak{p})^{\lfloor v_\mathfrak{p}(\mathfrak{n})/2\rfloor}\mathbf{1}_{v_\mathfrak{p}(x^2-4)\geqslant v_\mathfrak{p}(\mathfrak{n})}\right)dx\leqslant\begin{cases}1,&\mathfrak{p}\nmid 2,\\2|4|_\mathfrak{p}^{-1},&\mathfrak{p}\mid 2;\end{cases}$$

$$\int_{\mathfrak{o}_\mathfrak{p}}w_1^{\mathfrak{n},\mathfrak{p}}=\int_{\mathfrak{o}_\mathfrak{p}}N(\mathfrak{p})^{2v_\mathfrak{p}(\mathfrak{n})}\mathbf{1}_{v_\mathfrak{p}(x^2-4)\geqslant 2v_\mathfrak{p}(\mathfrak{n})}\,dx\leqslant\begin{cases}2,&\mathfrak{p}\nmid 2,\\4|4|_\mathfrak{p}^{-1},&\mathfrak{p}\mid 2.\end{cases}$$

Multiplying these local bounds over $\mathfrak{p}\mid\mathfrak{n}$, we get by Lemma 17 and (6.26) that

$$\int_{\widehat{\mathfrak{o}}}w_\kappa^\mathfrak{n}\leqslant 2^{\omega(\mathfrak{n})}8^{[k:\mathbb{Q}]}\preccurlyeq 1.\tag{6.28}$$

Combining (6.23), (6.27), (6.28), we see that $\operatorname{tr}Rf_\mathbb{A}\preccurlyeq\mathcal{C}(\Gamma_\kappa(\mathfrak{n}),\boldsymbol{T})$. Going back to (6.7), we have proved that

$$\sum_{\pi\in\mathcal{B}(\boldsymbol{\sigma},\boldsymbol{T})}\mathrm{m}(\pi,\Gamma_\kappa(\mathfrak{n}))\preccurlyeq e^{-\sum_{j\in S}2\rho_j R_j\sigma_j}\mathcal{C}(\Gamma_\kappa(\mathfrak{n}),\boldsymbol{T}).\tag{6.29}$$

This estimate holds for every choice of nonnegative parameters $\boldsymbol{R}=(R_j)_{j\in S}$ satisfying (6.17). If $S=\emptyset$, this proves Theorem 1. In fact, it is not hard to see that, for $S=\emptyset$, the proof gives (6.29) in the slightly stronger form

$$\sum_{\pi\in\mathcal{B}(\boldsymbol{\sigma},\boldsymbol{T})}\mathrm{m}(\pi,\Gamma_\kappa(\mathfrak{n}))\ll_\varepsilon\operatorname{vol}(\Gamma_\kappa(\mathfrak{n})\backslash G)^\varepsilon\mathcal{C}(\Gamma_\kappa(\mathfrak{n}),\boldsymbol{T}).\tag{6.30}$$

Otherwise, let $j_0\in S$ be such that σ_{j_0} is maximal, that is, $p(\boldsymbol{\sigma})=2/(1-2\sigma_{j_0})$ by (1.13). Then, by setting

$$R_j:=\begin{cases}\rho_{j_0}^{-1}(7+\log\mathcal{C}(\Gamma_\kappa(\mathfrak{n}),\boldsymbol{T})),&\text{if }j=j_0,\\0,&\text{if }j\neq j_0,\end{cases}\tag{6.31}$$

we get

$$\sum_{\pi \in \mathcal{B}(\boldsymbol{\sigma}, \boldsymbol{T})} \mathrm{m}(\pi, \Gamma_\kappa(\mathfrak{n})) \ll_\varepsilon \mathcal{C}(\Gamma_\kappa(\mathfrak{n}), \boldsymbol{T})^{1-2\sigma_{j_0}+\varepsilon} = \mathcal{C}(\Gamma_\kappa(\mathfrak{n}), \boldsymbol{T})^{2/p(\boldsymbol{\sigma})+\varepsilon}.$$

The proof of Theorem 1 is complete.

References

[1] M. Abért, N. Bergeron, I. Biringer, T. Gelander, N. Nikolov, J. Raimbault, I. Samet, *On the growth of L^2-invariants for sequences of lattices in Lie groups*, Ann. of Math. (2) **185** (2017), 711–790.

[2] J. Arthur, *An introduction to the trace formula*, In: Harmonic analysis, the trace formula, and Shimura varieties, 1–263, Clay Math. Proc., Vol. 4, Amer. Math. Soc., Providence, RI, 2005.

[3] V. Blomer, *Density theorems for* $\mathrm{GL}(n)$, Invent. Math. **232** (2023), 683–711.

[4] V. Blomer, J. Buttcane, P. Maga, *Applications of the Kuznetsov formula on* GL(3) *II: the level aspect*, Math. Ann. **369** (2017), 723–759.

[5] A. Borel, *Commensurability classes and volumes of hyperbolic* 3-*manifolds*, Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4) **8** (1981), 1–33.

[6] A. Borel, J. Tits, *Groupes réductifs*, Inst. Hautes Études Sci. Publ. Math. **27** (1965), 55–150.

[7] A. Borel, N. Wallach, *Continuous cohomology, discrete subgroups, and representations of reductive groups*, 2nd edition, Mathematical Surveys and Monographs, Vol. 67, Amer. Math. Soc., Providence, RI, 2000.

[8] R. W. Bruggeman, R. J. Miatello, *Sum formula for* SL_2 *over a number field and Selberg type estimate for exceptional eigenvalues*, Geom. Funct. Anal. **8** (1998), 627–655.

[9] F. Brumley, D. Milićević, *Counting cusp forms by analytic conductor*, Ann. Sci. Éc. Norm. Supér, to appear, arXiv:1805.00633

[10] J. W. S. Cassels, A. Fröhlich (eds.), *Algebraic number theory*, Academic Press, London, 1967.

[11] D. L. DeGeorge, N. R. Wallach, *Limit formulas for multiplicities in* $L^2(\Gamma \backslash G)$, Ann. of Math. (2) **107** (1978), 133–150.

[12] D. L. DeGeorge, N. R. Wallach, *Limit formulas for multiplicities in* $L^2(\Gamma \backslash G)$. *II. The tempered spectrum*, Ann. of Math. (2) **109** (1979), 477–495.

[13] A. Deitmar, S. Echterhoff, *Principles of harmonic analysis*, 2nd edition, Universitext, Springer, Cham, 2014.

[14] J. Dieudonné, *Treatise on analysis, Vol. VI.*, Translated from the French by I. G. Macdonald, Pure and Applied Mathematics, Vol. 10-VI, Academic Press, New York–London, 1978.

[15] J. Dixmier, C^*-*algebras*, Translated from the French by F. Jellett, North-Holland Mathematical Library, Vol. 15, North-Holland Publishing Co., Amsterdam-New York-Oxford, 1977.

[16] E. Dobrowolski, *On a question of Lehmer and the number of irreducible factors of a polynomial*, Acta Arith. **34** (1979), 391–401.

[17] T. Finis, E. Lapid, *An approximation principle for congruence subgroups II: application to the limit multiplicity problem*, Math. Z. **289** (2018), 1357–1380.

[18] T. Finis, E. Lapid, W. Müller, *Limit multiplicities for principal congruence subgroups of* $\mathrm{GL}(n)$ *and* $\mathrm{SL}(n)$, J. Inst. Math. Jussieu **14** (2015), 589–638.

[19] M. Frączyk, *Strong limit multiplicity for arithmetic hyperbolic surfaces and* 3-*manifolds*, Invent. Math. **224** (2021), 917–985.

[20] M. Frączyk, G. Harcos, P. Maga, *Counting bounded elements of a number field*, Int. Math. Res. Not. IMRN **2022**, no. 1, 373–390.

[21] P. Garrett, *Primer of spherical harmonic analysis on* $SL_2(\mathbb{C})$, http://www-users.math.umn.edu/~garrett/m/v/SL2C.pdf

[22] T. Gelander, *Homotopy type and volume of locally symmetric manifolds*, Duke Math. J. **124** (2004), 459–515.

[23] K. Golubev, A. Kamber, *Cutoff on graphs and the Sarnak–Xue density of eigenvalues*, European J. Combin. **104** (2022), Paper No. 103530, 23 pp.

[24] K. Golubev, A. Kamber, *On Sarnak's Density Conjecture and its applications*, Forum Math. Sigma **11** (2023), Paper No. e48, 51 pp.

[25] S. Helgason, *Groups and geometric analysis*, Corrected reprint of the 1984 original, Mathematical Surveys and Monographs, Vol. 83, Amer. Math. Soc., Providence, RI, 2000.

[26] R. Howe, C. Moore, *Asymptotic properties of unitary representations*, J. Functional Analysis **32** (1979), 72–96.

[27] P. Humphries, *Density theorems for exceptional eigenvalues for congruence subgroups*, Algebra Number Theory **12** (2018), 1581–1610.

[28] J. Huntley, Y. R. Katznelson, *Density theorems for congruence groups in real rank* 1, Duke Math. J. **71** (1993), 463–473.

[29] M. Huxley, *Exceptional eigenvalues and congruence subgroups*, In: The Selberg trace formula and related topics (Brunswick, Maine, 1984), 341–349, Contemp. Math., Vol. 53, Amer. Math. Soc., Providence, RI, 1986.

[30] H. Iwaniec, *Small eigenvalues of Laplacian for* $\Gamma_0(N)$, Acta Arith. **56** (1990), 65–82.

[31] H. Iwaniec, *Spectral methods of automorphic forms*, 2nd edition, Graduate Studies in Mathematics, Vol. 53, Amer. Math. Soc., Providence, RI, 2002.

[32] H. Iwaniec, E. Kowalski, *Analytic number theory*, Amer. Math. Soc. Colloquium Publications, Vol. 53, Amer. Math. Soc., Providence, RI, 2004.

[33] H. Iwaniec, P. Sarnak, *Perspectives on the analytic theory of L-functions*, GAFA 2000 (Tel Aviv, 1999), Geom. Funct. Anal. 2000, Special Volume, Part II, 705–741.

[34] S. Jana, *Applications of analytic newvectors for* GL(n), Math. Ann. **380** (2021), 915–952.

[35] J. Jorgenson, S. Lang, *The heat kernel and theta inversion on* SL$_2$(**C**), Springer Monographs in Mathematics, Springer, 2008.

[36] H. Kim, *Functoriality for the exterior square of* GL$_4$ *and the symmetric fourth of* GL$_2$, With Appendix 1 by D. Ramakrishnan and Appendix 2 by H. Kim and P. Sarnak, J. Amer. Math. Soc. **16** (2003), 139–183.

[37] A. Knapp, *Representation theory of semisimple groups. An overview based on examples*, Reprint of the 1986 original, Princeton Landmarks in Mathematics, Princeton University Press, Princeton, NJ, 2001.

[38] R. E. Kottwitz, *Harmonic analysis on reductive p-adic groups and Lie algebras*, In: Harmonic analysis, the trace formula, and Shimura varieties, 393–522, Clay Math. Proc., Vol. 4, Amer. Math. Soc., Providence, RI, 2005.

[39] S. Lang, SL$_2$(**R**), Reprint of the 1975 edition, Graduate Texts in Mathematics, Vol. 105, Springer-Verlag, New York, 1985.

[40] S. Lang, *Algebraic number theory*, 2nd ed., Graduate Texts in Mathematics, Vol. 110, Springer-Verlag, New York, 1994.

[41] C. MacLachlan, A. Reid, *The arithmetic of hyperbolic* 3-*manifolds*, Graduate Texts in Mathematics, Vol. 219, Springer, 2013.

[42] J. Matz, N. Templier, *Sato–Tate equidistribution for families of Hecke–Maass forms on* SL($n, \mathbb{R}$)/SO(n), Algebra Number Theory **15** (2021), 1343–1428.

[43] J. S. Milne, *Algebraic groups. The theory of group schemes of finite type over a field*, Cambridge Studies in Advanced Mathematics, Vol. 170, Cambridge University Press, Cambridge, 2017.

[44] J. Neukirch, *Algebraic number theory*, Grundlehren der Mathematischen Wissenschaften, Vol. 322, Springer-Verlag, Berlin, 1999.

[45] A. M. Odlyzko, *Some analytic estimates of class numbers and discriminants*, Invent. Math. **29** (1975), 275–286.

[46] T. Ono, *Arithmetic of algebraic tori*, Ann. of Math. (2) **74** (1961), 101–139.

[47] T. Ono, *On the Tamagawa number of algebraic tori*, Ann. of Math. (2) **78** (1963), 47–73.

[48] G. Prasad, *Galois-fixed points in the Bruhat–Tits building of a reductive group*, Bull. Soc. Math. France **129** (2001), 169–174.

[49] G. Rousseau, *Immeubles des groupes réductifs sur les corps locaux*, Ph.D. dissertation, Université de Paris-Sud Centre d'Orsay, 1977.

[50] P. Sarnak, *Statistical properties of eigenvalues of the Hecke operators*, In: Analytic number theory and Diophantine problems (Stillwater, OK, 1984), 321–331, Progr. Math., Vol. 70, Birkhäuser Boston, Boston, MA, 1987.

[51] P. Sarnak, *Diophantine problems and linear groups*, In: Proceedings of the International Congress of Mathematicians (Kyoto, 1990), 459–471, Math. Soc. Japan, Tokyo, 1991.

[52] P. Sarnak, *Notes on the generalized Ramanujan conjectures*, In: Harmonic analysis, the trace formula, and Shimura varieties, 659–685, Clay Math. Proc., Vol. 4, Amer. Math. Soc., Providence, RI, 2005.

[53] P. Sarnak, X. Xue, *Bounds for multiplicities of automorphic representations*, Duke Math. J. **64** (1991), 207–227.

[54] E. Ullmo, A. Yafaev, *Nombre de classes des tores de multiplication complexe et bornes inférieures pour les orbites galoisiennes de points spéciaux*, Bull. Soc. Math. France **143** (2015), 197–228.

[55] N. R. Wallach, *Real reductive groups II*, Pure and Applied Mathematics, Vol. 132-II, Academic Press, Boston, MA, 1992.

[56] A. Weil, *Basic number theory*, Springer-Verlag, New York, 3rd edition, 1974.

[57] A. Weil, *Adeles and algebraic groups*, With appendices by M. Demazure and T. Ono, Progress in Mathematics, Vol. 23, Birkhäuser, Boston, Mass., 1982.

[58] X. Xue, *On the Betti numbers of a hyperbolic manifold*, Geom. Funct. Anal. **2** (1992), 126–136.

ALFRÉD RÉNYI INSTITUTE OF MATHEMATICS, HUNGARIAN ACADEMY OF SCIENCES, POB 127, BUDAPEST H-1364, HUNGARY
Email address: `gharcos@renyi.hu, magapeter@gmail.com`

MTA RÉNYI INTÉZET LENDÜLET AUTOMORPHIC RESEARCH GROUP
Email address: `gharcos@renyi.hu, magapeter@gmail.com`

INSTITUTE FOR ADVANCED STUDY, 1 EINSTEIN DRIVE, PRINCETON, NJ 08540, USA
Current address: Department of Mathematics, University of Chicago, 5734 S University Ave, Chicago, IL 60637, USA
Email address: `mfraczyk@math.uchicago.edu`

BRYN MAWR COLLEGE, DEPARTMENT OF MATHEMATICS, 101 NORTH MERION AVENUE, BRYN MAWR, PA 19010, USA
Current address: Max-Planck-Institut für Mathematik, Vivatsgasse 7, D-53111 Bonn, Germany
Email address: `dmilicevic@brynmawr.edu`