Automorphic representations with prescribed ramification for unitary groups

William Conley[*]

May 2012

Abstract

Let F be a totally real number field, n a prime integer, and G a unitary group of rank n defined over F that is compact at infinity. We prove an asymptotic formula for the number of automorphic representations of G whose factors at finitely many places are prescribed up to inertia. These factors are specified by local inertial types, and the proof relies crucially on a bound on the traces of these types, which we establish first.

1 Introduction

In the representation theory of real and $\mathfrak{p}$-adic groups, much use is made of studying representations of a group G via their restrictions to certain compact subgroups. Here we use the same strategy for an adelic group G, in order to count the multiplicity of representations with certain ramification behavior in the automorphic spectrum of G. To specify such ramification behavior precisely, we will use the theory of types for $\mathfrak{p}$-adic groups, and will define a corresponding notion of global types for G. Our first goal is thus a purely local result, and the main theorem will follow as an application of this. We begin by describing the local work.

[*] Partially funded by EPSRC grant EP/G001480/1

1.1 Local Theory

For now, we let F be a nonarchimedean local field, and let G be the group of F-rational points of a connected reductive algebraic group defined over F. Let $\mathrm{Rep}(G)$ denote the category of smooth representations (with complex coefficients) of G, and let $\mathrm{Irr}(G)$ denote the set of isomorphism classes of irreducible objects in $\mathrm{Rep}(G)$. For any $\pi \in \mathrm{Irr}(G)$, there exist a parabolic subgroup P of G with Levi subgroup L, and an irreducible supercuspidal representation σ of L, such that π is a subquotient of $\mathrm{ind}_P^G(\sigma)$. (Here ind_P^G denotes the functor of normalized parabolic induction.) The pair (L, σ) is uniquely determined by π up to conjugation by an element of G, and is referred to as the supercuspidal support of π. Let $[L, \sigma]_G$ denote the equivalence class of the pair (L, σ) modulo twisting by unramified characters of L and conjugation by elements of G. Then $[L, \sigma]_G$ is called the inertial support of π. Let $\mathcal{B}(G)$ be the set of all inertial supports of representations in $\mathrm{Irr}(G)$.

For $\mathfrak{s} = [L, \sigma]_G \in \mathcal{B}(G)$, let $\mathrm{Rep}^{\mathfrak{s}}(G)$ denote the full subcategory of $\mathrm{Rep}(G)$ consisting of representations whose irreducible subquotients all have inertial support $\mathfrak{s}$. In other words, the irreducible objects in $\mathrm{Rep}^{\mathfrak{s}}(G)$ are precisely the irreducible subquotients of $\mathrm{ind}_P^G(\chi\sigma)$, as χ runs through the set of all unramified characters of L. A *type* for $\mathfrak{s}$ is a pair (J, λ) consisting of a compact open subgroup J of G and an irreducible representation λ of J such that, for all $\pi \in \mathrm{Irr}(G)$,

$$\mathrm{Hom}_J(\lambda, \pi) \neq 0 \iff \pi \in \mathrm{Rep}^{\mathfrak{s}}(G).$$

When $\mathrm{Hom}_J(\lambda, \pi) \neq 0$, we will simply say that π contains λ. While this notion of type has been well developed for a large class of groups and has many interesting applications, for our purposes, it is inconvenient to allow the subgroup J to vary. Instead, we will work with a fixed maximal compact subgroup K of G. However, it is not possible to define types on K for all $\mathfrak{s} \in \mathcal{B}(G)$. For example, let $G = \mathrm{GL}_n(F)$, $n \geq 2$, and $\mathfrak{s}_0 = [T, 1_T]_G$, where T is the diagonal subgroup of G and 1_T denotes the trivial character. Then $\mathrm{Rep}^{\mathfrak{s}_0}(G)$ contains the unramified principal series and the Steinberg representations of G, among others. It is well known that, letting I denote an Iwahori subgroup of G, the pair $(I, 1_I)$ is a type for $\mathfrak{s}_0$. However, there is no type for $\mathfrak{s}_0$ of the form (K, λ).

Therefore, we will work with the following variant of the theory of types, introduced by Henniart in [?]. For $\mathfrak{s} \in \mathcal{B}(G)$, we will say that $\tau \in \mathrm{Irr}(K)$ is *typical* for $\mathfrak{s}$ if, for all $\pi \in \mathrm{Irr}(G)$,

$$\mathrm{Hom}_K(\tau, \pi) \neq 0 \implies \pi \in \mathrm{Rep}^{\mathfrak{s}}(G).$$

Note that by Frobenius reciprocity, if (J, λ) is a type for $\mathfrak{s}$ with $J \subset K$, then the irreducible components of $\mathrm{Ind}_J^K(\lambda)$ are all typical for $\mathfrak{s}$, and every irreducible representation $\pi \in \mathrm{Rep}^{\mathfrak{s}}(G)$ contains at least one of these irreducible components. Suppose now that there exist types for all $\mathfrak{s} \in \mathcal{B}(G)$. (This was proved for $\mathrm{GL}_n(F)$ in [**?**], the case in which we are most interested here. More generally, this is believed to be true for all reductive groups G, but is currently known only for a few classes of groups beyond the general linear ones.) Fix representatives $K_1, \ldots, K_r$ of the conjugacy classes of maximal compact subgroups of G. Then by the observation above, we immediately have the following: For each $\mathfrak{s} \in \mathcal{B}(G)$ and each irreducible representation $\pi \in \mathrm{Rep}^{\mathfrak{s}}(G)$, there exists (at least one) $i \in \{1, \ldots, r\}$ for which there exists (at least one) $\tau \in \mathrm{Irr}(K_i)$ such that τ is typical for $\mathfrak{s}$ and π contains τ.

Extending the previous example of $G = \mathrm{GL}_n(F)$ and $\mathfrak{s}_0 \in \mathcal{B}(G)$, we fix $K = \mathrm{GL}_n(\mathfrak{o}_F)$ and let I be the standard Iwahori subgroup in K. Then since 1_K is an irreducible component of $\mathrm{Ind}_I^K(1_I)$, we see that 1_K is typical for $\mathfrak{s}_0$. Indeed, it is well known that the irreducible representations of G whose restriction to K contains a copy of the trivial character are precisely the unramified ones. (In fact, this holds for any reductive group G over F, and is often taken as the definition of an unramified representation. This fact will play a crucial role in our definition of global types below.) Returning to $G = \mathrm{GL}_n(F)$, the Steinberg representation of G does not contain 1_K, so that 1_K is not a type for $\mathfrak{s}_0$. The inflation of the Steinberg representation of $\mathrm{GL}_n(\boldsymbol{k}_F)$ to K is another irreducible component of $\mathrm{Ind}_I^K(1_I)$, and hence is also typical for $\mathfrak{s}_0$. This representation is contained in the unramified principal series and the Steinberg representations of G, but not, for example, in the one-dimensional unramified representations of G.

Our first result is to establish, for a large class of fixed elements g of K, a bound on the trace of $\tau(g)$ as τ runs over all typical representations for the supercuspidal components of $\mathrm{GL}_n(F)$, where n is assumed to be prime. This is the content of Theorem 3.1 below, and poses the most significant hurdle to the proof of our main theorem. It is likely that a similar trace bound holds in much greater generality, in particular without the restrictions that n be prime, and for all typical representations of K rather than just the supercuspidals. While we use this result here primarily as a tool to prove our main global theorem, this purely local result is rather interesting in its own right.

1.2 Global Theory

We now shift our focus to the global setting, and correspondingly shift our notation as well. So in this section, F will denote a number field, and G will denote a connected reductive algebraic group defined over F. Let Z be the center of G.

Let $G(\mathbb{A})$ denote the group of adelic points of G. For each finite place v of F, fix a maximal compact subgroup K_v of $G(F_v)$, and for each infinite place v, let $K_v = G(F_v)$. Let $K = \prod_v K_v \subset G(\mathbb{A})$, and let Z_0 be the center of K.

Suppose now that $\pi = \bigotimes'_v \pi_v$ is an automorphic representation of $G(\mathbb{A})$ such that π_v contains some typical representation τ_v defined on K_v at every finite place v. Since π_v is unramified at almost all finite places, we may take τ_v to be the trivial representation of K_v for almost all v. We define $\tau_v = \pi_v$ for all $v \mid \infty$. Then since π is trivial on $G(F)$, and thus its central character ω_π is trivial on $Z(F)$, it must be true that $\prod_v \omega_{\tau_v}$ is trivial on $Z_0 \cap G(F)$. Motivated by this, we make the following definition.

Definition 1.1. A *global type* on K for the group G is a representation $\tau = \bigotimes_v \tau_v$ of K satisfying the following:

1. For all $v \nmid \infty$, τ_v is an irreducible representation of K_v that is typical, and for almost all of these, τ_v is trivial.

2. For all $v \mid \infty$, τ_v is an irreducible representation of $K_v = G(F_v)$.

3. The central character $\omega_\tau = \prod_v \omega_{\tau_v}$ is trivial on $Z_0 \cap G(F)$.

(Note that because of the first condition here, both products appearing in this definition are actually finite.)

It should be clear now that global types are defined to encode information about the local factors of an automorphic representation at every place. Given a global type τ on K, we will of course say that an automorphic representation π of G contains τ, or is of type τ, if $\pi_v \cong \tau_v$ for all $v \mid \infty$ and π_v contains τ_v for all $v \nmid \infty$. Thus at the infinite places, a global type specifies π_v completely; at the finite places, it describes π_v up to inertia.

Let $\mathcal{A}(\,G(F)\backslash G(\mathbb{A})\,)$ denote the space of all automorphic forms on $G(\mathbb{A})$, of which the automorphic representations of $G(\mathbb{A})$ are the irreducible subquotients. For any global type τ, let $m(\tau)$ be the multiplicity of τ in the restriction of $\mathcal{A}(\,G(F)\backslash G(\mathbb{A})\,)$ to K, and let $\mathcal{R}(\tau)$ denote the set of distinct isomorphism classes of automorphic representations of $G(\mathbb{A})$ of type τ. Our goal is to compute bounds on the cardinality of $\mathcal{R}(\tau)$, for a certain class of groups G and global types τ. We will do so by computing bounds on $m(\tau)$. Note, however, that in general we cannot directly relate $m(\tau)$ to $\#\mathcal{R}(\tau)$, for two reasons: First, an automorphic representation π may occur with multiplicity greater than 1 in $\mathcal{A}(\,G(F)\backslash G(\mathbb{A})\,)$ (i.e., multiplicity one may fail to be true for G.) Second, a global type τ may occur with multiplicity greater than 1 in the restriction of an automorphic representation π to K. If we restrict our attention to a class of global types for which the latter phenomenon does not occur, as we will below, then we may conclude that $m(\tau) \geq \#\mathcal{R}(\tau)$. On the other hand, if multiplicity one holds for G, then $m(\tau) \leq \#\mathcal{R}(\tau)$. Thus only if both conditions hold are we guaranteed an equality. In any case, however, we certainly have that $\mathcal{R}(\tau)$ is nonempty if and only if $m(\tau) \neq 0$.

When F is a totally real field, $G = \mathrm{GL}_2$, and one considers global types that are discrete series with weights of equal parity at all of the infinite places, then the automorphic representations under consideration correspond to Hilbert modular cusp forms. In [?], Weinstein considered this case, and obtained bounds on the size of $\mathcal{R}(\tau)$ for all such global types. In particular, his result showed that, with the exception of a finite number of twist classes of types, such automorphic representations do exist. Generalizing such a result beyond GL_2 presents a number of immediate difficulties:

1. For $G = \mathrm{GL}_n$ with $n > 2$, there are no discrete series at infinity. Many of the global methods used to count automorphic representations are best suited to those whose infinity type is discrete.

2. The theory of types has not been completely developed for many groups other than GL_n. Furthermore, even for GL_n, the general theory quickly becomes extremely complicated.

3. The theory of typical representations defined on a maximal compact subgroup, outlined above, has not even been fully developed for GL_n, $n > 2$, beyond the supercuspidal case.

As a result of these complications, we have chosen here to work in a rather special setting. First, we will work with unitary groups, for a number of reasons. For one thing, unitary groups always have discrete series at infinity. Furthermore, this choice allows us to use the rather well developed theory of types for GL_n at the split places (i.e., where $G(F_v) \cong \mathrm{GL}_n(F_v)$). And finally, tremendous progress has been made in recent years on global functoriality for unitary groups, relating automorphic representations of unitary groups to those of other groups, as well as to global Galois representations. Thus, the results we obtain should have applications to other settings. We now describe more precisely and justify briefly the restrictions that we will make:

1. We will choose G to be a unitary group, defined over a totally real field F, relative to a totally imaginary quadratic extension E of F. Furthermore, in order to greatly simplify the global methods used, we will work with a unitary group that is compact at infinity (i.e., such that $G(F_v) \cong \mathrm{U}(n)$ for all $v \mid \infty$).

2. Because the theory of typical representations is undeveloped for unitary groups, and is furthermore only well understood for supercuspidal representations of GL_n, we will restrict our attention to global types that are (up to twisting) trivial at all of the non-split places, and either trivial or supercuspidal at all of the split places. Let $\mathcal{T}_G$ denote the set of all such global types for G.

3. Finally, because the theory of types for supercuspidal representations of GL_n becomes vastly more complicated when n is composite, we will restrict our attention here to the case where n is prime. (That is, we will choose G to have prime rank.)

For the complete details, see section 5 below. With these restrictions in mind, we may now state our main theorem. For a global type τ on a unitary group G as described above, let $S(\tau)$ be the set of finite places v for which $\dim(\tau_v) > 1$. To deal with the infinite places, let $\mathfrak{h}_{\mathbb{R}}^*$ denote the space of weight vectors for the Lie group $\mathrm{U}(n)$. (See section 4 below for details.) For a global type τ and an infinite place v of F, we will denote by $\lambda_v(\tau) \in \mathfrak{h}_{\mathbb{R}}^*$ the highest weight vector of τ_v (viewed as a representation of $\mathrm{U}(n)$). The Weyl dimension formula then gives the dimension of τ_v as a polynomial function of $\lambda_v(\tau)$, of degree $\frac{n^2-n}{2}$. We will refer to this polynomial as the Weyl polynomial of $\mathrm{U}(n)$. Our main theorem is now

Theorem 1.2. *For all global types $\tau \in \mathcal{T}_G$,*

$$m(\tau) = \#\mu_E \cdot \mathfrak{m}(G,K) \cdot \dim(\tau) + O\left(n^{\#S(\tau)} \cdot \prod_{v|\infty} P_v(\lambda_v(\tau)) \right),$$

where μ_E is the group of roots of unity of E, $\mathfrak{m}(G,K)$ is the mass of G relative to K as in [?], and the P_v are polynomials on $\mathfrak{h}_{\mathbb{R}}^*$ of degree strictly less than that of the Weyl polynomial of $\mathrm{U}(n)$.

Of course, the polynomials P_v and the constant implicit in the O in this formula depend only on the group G. We give a formula for $\mathfrak{m}(G,K)$ in Section 6 below. For any particular example of a group G, the other terms in the formula above could be computed explicitly as well, so that the theorem could yield quite precise results. But we note that in any event, we have the following immediate corollary.

Corollary 1.3. *For all but a finite number of twist classes of global types $\tau \in \mathcal{T}_G$, there exist automorphic representations of $G(\mathbb{A})$ of type τ.*

Proof. For $v \in S$, the smallest possible dimension of a supercuspidal type defined on $K_v \cong \mathrm{GL}_n(\mathfrak{o}_{F_v})$ is $(q_v - 1)(q_v^2 - 1) \cdots (q_v^{n-1} - 1)$, where q_v is the cardinality of the residue field of F_v. (See section 3 below for details.) Obviously this is greater than n for almost all v. Let P denote the Weyl polynomial for $\mathrm{U}(n)$. Then for any infinite place v of F, $\dim(\tau_v) = P(\lambda_v(\tau))$, so

$$\dim(\tau) \geq \prod_{v \in S(\tau)} \left((q_v - 1) \cdots (q_v^{n-1} - 1) \right) \cdot \prod_{v|\infty} P(\lambda_v(\tau))$$

for all $\tau \in \mathcal{T}_G$. Thus, assuming the notation of Theorem 1.2, if we enumerate the twist classes of global types $\tau \in \mathcal{T}_G$, it is clear that

$$\frac{\dim(\tau)}{n^{\#S(\tau)} \cdot \prod_{v|\infty} P_v(\lambda_v(\tau))}$$

grows without bound. The result now follows. □

1.3 Remarks

Our main theorem and the methods used to prove it are a generalization of those in [?]. In particular, as mentioned previously, the key ingredient is to establish, for a large class of elements $g_v \in K_v$, a bound on the trace of $\tau_v(g_v)$ as τ varies over all global types in $\mathcal{T}_G$. Proving these bounds at the finite places is the most significant obstacle. We begin by recalling the construction of types for the supercuspidal representations of $\mathrm{GL}_n(F)$, in the relatively straightforward case where n is prime, in Section 2. We prove the trace bound on such types in Section 3 (Theorem 3.1). There we once again make extensive use of the fact that n is assumed to be prime. In Section 4, we review some of the representation theory of $\mathrm{U}(n)$ that we will need in order to deal with the infinite places. We give a precise description of the group and the global types that we will be considering in Section 5. Finally, in Section 6, we prove Theorem 1.2.

There are a number of directions in which this result could be generalized. Perhaps the most obvious would be to remove the restriction that n be prime. It is likely that a statement similar to Theorem 3.1 is true without this restriction, but it is clear even from the depth zero case that the hypotheses would need to be strengthened in some way. Furthermore, as mentioned above, the full construction of types for supercuspidal representations of $\mathrm{GL}_n(F)$, for arbitrary n, is significantly more complicated than when n is prime. In particular, the groups H^1, J^1, and J (see Section 2.1) used in the definition must be specified in terms of defining sequences for the underlying simple strata (see [**?**] for details). The approach used here for the positive depth cases seems insufficient to deal with this added complexity.

Another direction in which to take this would be to use types for p-adic unitary groups at the non-split places, rather than only considering global types that are trivial there (and thus restricting the result to automorphic representations that are unramified at these places). Types for the supercuspidal representations of p-adic unitary groups have been constructed in [**?**]. Similarly, it should be possible to treat more cases than just representations that are supercuspidal or twists of an unramified principal series, using a more general class of types than the ones used here. Again, it is likely that a version of Theorem 3.1 will remain true for such types, but two complications arise: the construction of such types becomes further complicated and breaks up into many special cases; and there are representations of G that contain multiple typical representations of K, often with multiplicities greater than one.

Finally, the restriction that the global unitary group be compact at all infinite places greatly simplifies the global argument used here, but it is likely that an application of the trace formula to global types like the ones constructed here would yield similar results for a much greater class of groups. It is also worth noting that Shin has recently announced, in [?], a result along similar lines. While Shin's result applies to a much more general class of groups and automorphic representations, the asymptotic formula he derives, specialized to this case, is quite different from ours. The reason for this is that we are counting the raw number of automorphic representations of a given type, whereas his formula estimates the total dimension of certain isotypic subspaces within such representations.

Acknowledgements

I must first and foremost thank Jared Weinstein, on whose original work this paper is based. Great thanks are also due to Don Blasius for his guidance, and for looking over an earlier draft of this work. Without either of them, this paper would never have come to be. I would also like to thank Shaun Stevens for pointing out a mistake in an earlier draft of this work, and for many other helpful suggestions.

2 Maximal simple types when n is prime

In this section and the next, we will focus on the local theory at the nonarchimedean places, so we return to the notation of section 1.1. Namely, let F be a nonarchimedean local field, with ring of integers $\mathfrak{o}_F$, prime ideal $\mathfrak{p}_F$, and residue field $\boldsymbol{k}_F = \mathfrak{o}_F/\mathfrak{p}_F$ of cardinality q. Throughout this entire section, we will assume that n is prime. Though the types defined below were first constructed by Carayol in [?], our description of them follows [?] exactly, as does our basic notation and terminology.

2.1 Definition of types

As we will occasionally make use of explicit matrix computations, we fix $V = F^n$, and we fix a basis of V so that we may identify $A = \mathrm{End}_F(V)$ with $M_n(F)$ and $G = A^\times$ with $\mathrm{GL}_n(F)$. We also fix a choice of additive character ψ of F of level zero. The construction of types in this setting breaks up naturally into three cases:

The depth zero case

Let τ be the twist by a character of $\mathfrak{o}_F^\times$ of the inflation to K of a cuspidal irreducible representation of $\mathrm{GL}_n(\boldsymbol{k}_F)$. Then τ is a supercuspidal type for G.

The unramified case

Let $\mathfrak{A} = M_n(\mathfrak{o}_F)$, a hereditary $\mathfrak{o}_F$-order in A, and let $\mathfrak{P} = \mathfrak{p}M_n(\mathfrak{o}_F)$, the Jacobson radical of $\mathfrak{A}$. Let $\beta \in A \setminus \mathfrak{A}$ such that $E = F[\beta]$ is an unramified field extension of F of degree n, such that $E^\times$ normalizes $\mathfrak{A}$, and such that β is *minimal* over F (see [**?**, 1.4.14]). Let m be the unique (positive) integer such that $\beta \in \mathfrak{P}^{-m} \setminus \mathfrak{P}^{-m+1}$. Define a character ψ_β of the group $1 + \mathfrak{P}^{\lceil \frac{m+1}{2} \rceil}$ by

$$\psi_\beta(x) = \psi(\mathrm{Tr}_{A/F}(\beta(x-1))).$$

Define groups H^1, J^1, and J as follows:

$$\begin{aligned} H^1 &= (1+\mathfrak{p}_E)(1+\mathfrak{P}^{\lceil \frac{m+1}{2} \rceil}), \\ J^1 &= (1+\mathfrak{p}_E)(1+\mathfrak{P}^{\lfloor \frac{m+1}{2} \rfloor}), \text{ and} \\ J &= \mathfrak{o}_E^\times(1+\mathfrak{P}^{\lfloor \frac{m+1}{2} \rfloor}). \end{aligned}$$

Let θ be any extension of ψ_β to H^1. There is a unique irreducible representation η of J^1 whose restriction to H^1 contains θ. Let λ be any extension of η from J^1 to J. The pair (J, λ) is now a special case of maximal simple type, in the language of [**?**]. Finally, let $\tau = (\chi \circ \det) \otimes \mathrm{Ind}_J^K(\lambda)$, for any character χ of $\mathfrak{o}_F^\times$. Then τ is a supercuspidal type for G.

The ramified case

Let $\mathfrak{A}$ be the subring of matrices in $M_n(\mathfrak{o}_F)$ that are upper triangular modulo $\mathfrak{p}$, which is also a hereditary $\mathfrak{o}_F$-order in A. Let $\mathfrak{P}$ again be the Jacobson radical of $\mathfrak{A}$, which is the ideal of matrices whose reductions modulo $\mathfrak{p}$ are nilpotent upper triangular. Much like before, let $\beta \in A \setminus \mathfrak{A}$ such that $E = F[\beta]$ is a totally ramified field extension of F of degree n, such that $E^\times$ normalizes $\mathfrak{A}$, and such that β is minimal over F. Define m, the groups H^1, J^1, and J, the characters ψ_β and θ, and the representations η, λ, and τ exactly as in the unramified case above. Once again, (J, λ) is a special case of maximal simple type, and τ is likewise a supercuspidal type for G.

The main result of [**?**] is that every supercuspidal type for G that is defined on K is one of the representations τ described above. Furthermore, for any irreducible supercuspidal representation π of G, the restriction of π to K contains one and only one such type, and that type occurs with multiplicity one in π.

2.2 A preliminary trace bound

Our first task is to more carefully analyze the representation λ in one particular case, namely when E is unramified and λ is *not* 1-dimensional (i.e., when m is even). The result that we derive here is probably well known to the experts, but the exact statement that we require does not seem to appear in the literature. At any rate, the details are quite technical, so we collect them here. It is likely that a very similar statement holds more generally, but the lemma below is sufficient for our needs. The proof of this lemma is very similar to others found in the literature (see for example [**?**, 4.1 - 4.2], [**?**, 4.1]), but adapted to the current setting.

Lemma 2.1. *Let $\mathfrak{A}$, β, and m be as in the unramified case described above, and assume that m is even. Let H^1, J^1, J, θ, η, and λ also be as above. Then*

$$|\operatorname{Tr}\lambda(a(1+x))| = 1$$

for any $x \in \mathfrak{P}^{\lfloor\frac{m+1}{2}\rfloor}$ and any $a \in \mathfrak{o}_E^\times$ whose reduction modulo $\mathfrak{p}_E$ is not in $\boldsymbol{k}_F^\times$.

Proof. For convenience, let $k = \left\lfloor\frac{m+1}{2}\right\rfloor = \frac{m}{2}$, so that

$$\begin{aligned} H^1 &= (1+\mathfrak{p}_E)(1+\mathfrak{P}^{k+1}), \\ J^1 &= (1+\mathfrak{p}_E)(1+\mathfrak{P}^{k}), \text{ and} \\ J &= \mathfrak{o}_E^\times(1+\mathfrak{P}^{k}). \end{aligned}$$

Note that we have an exact sequence

$$1 \to J^1 \to J \to \boldsymbol{k}_E^\times \to 1, \tag{2.1}$$

which in this case splits since $\boldsymbol{k}_E^\times \cong \mu_E$, the group of roots of unity of order prime to p in E. Thus $J = \boldsymbol{k}_E^\times \ltimes J^1$, where the action of $\boldsymbol{k}_E^\times$ on J^1 is by conjugation.

Recall (from [**?**, 3.3.1] for example) that θ is fixed under conjugation by J. Thus $\operatorname{Ker}\theta \lhd J$, so we let

$$\overline{H^1} = H^1\big/\operatorname{Ker}\theta\ , \quad \overline{J^1} = J^1\big/\operatorname{Ker}\theta\ , \quad \overline{J} = J/\operatorname{Ker}\theta\ ,$$

and let $\overline{\theta}$ (resp. $\overline{\eta}$, $\overline{\lambda}$) be the composition of θ (resp. η, λ) with the quotient map. Thus $\overline{\eta}$ is the unique irreducible representation of $\overline{J^1}$ whose restriction to $\overline{H^1}$ contains $\overline{\theta}$, and $\overline{\lambda}$ is an extension of $\overline{\eta}$ to $\overline{J}$.

Let $W = \overline{J^1}\Big/\overline{H^1} \cong J^1/H^1$, and define $h_\theta : W \times W \to \mathbb{C}^\times$ by

$$(\overline{x}, \overline{y}) \mapsto \theta[x, y].$$

By [**?**, 3.4.1], h_θ is a nondegenerate alternating bilinear form on the $\boldsymbol{k}_F$-vector space W, from which it follows that $\overline{H^1}$ is the center of $\overline{J^1}$ (and hence also of $\overline{J}$).

Although we will not need this result, note that in this setting

$$W \cong (1+\mathfrak{P}^k)/(1+\mathfrak{p}_E^k)(1+\mathfrak{P}^{k+1}) \cong \mathfrak{P}^k/\mathfrak{p}_E^k + \mathfrak{P}^{k+1}$$

is a $\boldsymbol{k}_F$-vector space of dimension n^2-n. Thus the representation λ will have dimension $q^{\frac{n^2-n}{2}}$.

The split exact sequence (2.1) reduces to

$$1 \to \overline{J^1} \to \overline{J} \to \boldsymbol{k}_E^\times \to 1,$$

which still splits. We regard $\boldsymbol{k}_E^\times$ as a group of automorphisms of $\overline{J^1}$, acting by conjugation. Fix $a \in \boldsymbol{k}_E^\times \smallsetminus \boldsymbol{k}_F^\times$. Note that the commutator map $W \to W$ defined by $v \mapsto a^{-1}vav^{-1}$ is an isomorphism. Thus if $g \in \overline{J^1}$, we can choose $g_0 \in \overline{J^1}$ such that $a^{-1}g_0ag_0^{-1} = gh^{-1}$ for some $h \in \overline{H^1}$, whence $g_0^{-1}(ag)g_0 = ah$ since $\overline{H^1}$ is the center of $\overline{J^1}$. Thus every element of $\overline{J}$ of the form ag, $g \in \overline{J^1}$, is conjugate to an element of the form ah, with $h \in \overline{H^1}$. So we will be finished if we can prove that $\left|\operatorname{Tr}\overline{\lambda}(ah)\right| = 1$ for all $h \in \overline{H^1}$.

Let $A = \langle a \rangle \subset \boldsymbol{k}_E^\times$. Note that $\overline{J^1}$ is a finite p-group (where p is the characteristic of $\boldsymbol{k}_F$), so its order is relatively prime to that of A. Since $\overline{\theta}$ is fixed by the action of A, the isomorphism class of $\overline{\eta}$ is as well. Under these circumstances, in [**?**], Glauberman gives a one-to-one correspondence between isomorphism classes of irreducible representations of $\overline{J^1}$ fixed by A and those of $\overline{J^1}^A = \overline{H^1}$. This correspondence maps $\overline{\eta}$ to $\overline{\theta}$ (by Theorem 5(d) of [**?**], for example). By Theorem 2 of [**?**], there exists a certain canonical extension of $\overline{\eta}$ to $\overline{J}$, and $\overline{\lambda}$ is a twist of it by a uniquely determined character χ of $\boldsymbol{k}_E^\times$. Thus by Theorem 3 of [?], there exists a constant $\epsilon = \pm 1$ such that

$$\operatorname{Tr} \overline{\lambda}(ah) = \epsilon\, \chi(a) \overline{\theta}(h)$$

for all $h \in \overline{H^1}$. (Note that the constant ϵ depends on a and on η, but this need not concern us here.) The result now follows. □

3 A bound on the characters of types

We now come to the first real result of this article. This is our main local result, and will provide the crucial ingredient in the proof of the main theorem.

Theorem 3.1. *Let n be a prime integer, let $g \in K = \mathrm{GL}_n(\mathfrak{o}_F)$, and assume g is not in the center of K. There exists a constant C_g such that for all supercuspidal types τ defined on K,*

$$|\mathrm{Tr}(\tau(g))| \leq C_g.$$

Let $\overline{g} \in \mathrm{GL}_n(\boldsymbol{k}_F)$ be the reduction of g modulo $\mathfrak{p}_F$. Then if the characteristic polynomial of $\overline{g}$ is irreducible, we may take $C_g = n$. Otherwise, if $\overline{g}$ has at least two distinct eigenvalues, then we may take $C_g = 0$.

Proof. We prove this theorem in three cases, corresponding to the three cases in the construction of types described in section 2.1. Since twisting clearly has no bearing on the conclusions stated here, we may ignore the twisting by characters of $\mathfrak{o}_F^\times$ that occurs as the last step of each of those cases.

3.1 The depth zero case

Let τ be a depth zero type. Then after twisting by a character of $\mathfrak{o}_F^\times$, we may assume that τ is merely the inflation to K of an irreducible cuspidal representation of $\mathrm{GL}_n(\boldsymbol{k}_F)$. Since this group is finite, the first claim is clear for this case.

The characters of the irreducible cuspidal representations of $\mathrm{GL}_n(\mathbb{F}_q)$ were first computed in [?]. We briefly recall the resulting formula, in a simplified form. Let $\mathbb{F}_{q^n}$ denote the finite field of q^n elements. A character θ of $\mathbb{F}_{q^n}^\times$ is called *regular* if its orbit under the action of $\mathrm{Gal}(\mathbb{F}_{q^n}/\mathbb{F}_q)$ has exactly n

elements, or in other words, if $\theta, \theta^q, \ldots, \theta^{q^{n-1}}$ are all distinct. The cuspidal representations of $\mathrm{GL}_n(\mathbb{F}_q)$ are in one-to-one correspondence with the orbits of these characters. Thus, for a regular character θ, we will denote the corresponding irreducible cuspidal representation of $\mathrm{GL}_n(\mathbb{F}_q)$ by τ_θ. Then [**?**, p. 431] gives the following:

- If the characteristic polynomial of $\overline{g}$ is a power of a single irreducible polynomial f of degree d, then
$$\mathrm{Tr}(\tau_\theta(\overline{g})) = (-1)^{n-1} \left(\prod_{i=1}^{r-1} (1 - q^i) \right) \left(\sum_\gamma \theta(\gamma) \right), \tag{3.1}$$
where the sum is taken over the d distinct roots γ of f in $\mathbb{F}_{q^n}$, and r is the number of Jordan blocks in the Jordan normal form of $\overline{g}$ over $\mathbb{F}_q$.
- Otherwise, $\mathrm{Tr}(\tau_\theta(\overline{g})) = 0$.

Note that the sum in (3.1) is invariant under the action of $\mathrm{Gal}(\mathbb{F}_{q^n}/\mathbb{F}_q)$, and thus depends only on the orbit of θ, as required. Since we are assuming n is prime, the first of these cases can happen only if the characteristic polynomial of $\overline{g}$ is either irreducible or of the form $(x - \gamma)^n$ for some $\gamma \in \mathbb{F}_q^\times$. Thus, if $\overline{g}$ has irreducible characteristic polynomial, we get $|\mathrm{Tr}(\tau_\theta(\overline{g}))| \leq n$ as desired. And otherwise, if $\overline{g}$ has at least two distinct eigenvalues, then $\mathrm{Tr}(\tau_\theta(\overline{g})) = 0$. Furthermore, taking $\overline{g} = 1$, we see that the dimension of τ_θ, and thus of any depth zero type, is $(q-1)(q^2-1)\cdots(q^{n-1}-1)$.

3.2 The unramified case

Let $\mathfrak{A} = M_n(\mathfrak{o}_F)$, and temporarily assume all of the other notation from the unramified case of section 2.1. Recall that in this case, $E = F[\beta]$ is an unramified extension of F of degree n, where $\beta \in M_n(F)$ is minimal over F. (Note also that since $\beta \in M_n(F)$, we regard E as being explicitly embedded in the F-algebra $M_n(F)$.) To simplify notation, we let $k = \left\lfloor \frac{m+1}{2} \right\rfloor$, so that $J = \mathfrak{o}_E^\times(1 + \mathfrak{P}^k)$. Since the final step in the construction of the type τ in this case was induction from J to K, we will naturally make use of the Frobenius formula:
$$\mathrm{Tr}\,\mathrm{Ind}_J^K(\lambda)(g) = \sum_{\substack{x \in K/J \\ x^{-1}gx \in J}} \mathrm{Tr}\,\lambda(x^{-1}gx). \tag{3.2}$$

Note that in this formula, the condition $x^{-1}gx \in J$ is equivalent to $gxJ = xJ$, or in other words that the coset of x in K/J is fixed under the natural left action of K. Thus to apply this formula, we will begin by defining a model of the coset space K/J that is equipped with the same left action of K, then determine the points fixed by the element g in this space.

As a starting point for our model of this coset space, note that there is a natural left action of $\mathrm{GL}_n(F)$ on $\mathbb{P}^{n-1}(E)$. (If we think of elements of projective space as column vectors in homogeneous coordinates, this action is just given by matrix multiplication.) Note that the subset of elements whose homogeneous coordinates form a basis of E over F is stable under this action, and the group acts transitively on this set. Similarly, we have a natural left action of K on $\mathbb{P}^{n-1}(\mathfrak{o}_E)$, and we define X to be the set of all points in $\mathbb{P}^{n-1}(\mathfrak{o}_E)$ with homogeneous coordinates

$$[u_0 : \ldots : u_{n-1}]$$

such that $\{u_0, \ldots, u_{n-1}\}$ is an $\mathfrak{o}_F$-basis of $\mathfrak{o}_E$. It is clear that K acts transitively on X, and that $\mathfrak{o}_E^\times$ is the stabilizer of some point $x \in X$. A simple computation shows that the action of the normal subgroup $1 + \mathfrak{P}^k$ induces the equivalence relation of congruence modulo $\mathfrak{p}_E^k$ on the coordinates u_i of points in X. If we let X_k denote the quotient of X under this equivalence (which we may think of as a subset of $\mathbb{P}^{n-1}(\mathfrak{o}_E/\mathfrak{p}_E^k)$), and let x_k denote the class of x, then we have a K-equivariant bijection

$$K/J \to X_k$$

defined by $aJ \mapsto a \cdot x_k$.

Note that the choice of the point x will depend on the embedding of E into $M_n(F)$, and hence on the element $\beta \in M_n(F)$. Thus the actual bijection established here will vary for different types (J, λ), even for different ones having the same value of k. However, the action of K on X_k is the same in all cases, and it will turn out that this will be all that matters for our purpose: since λ has dimension either 1 or $q^{\frac{n^2-n}{2}}$ (depending on whether m is odd or even, respectively), $\mathrm{Tr}(\lambda(g))$ is bounded by the latter value, so by the Frobenius formula, $\mathrm{Tr}(\tau(g))$ is bounded by this value times the number of fixed points of g in X_k. Thus the first claim of the theorem will be proved for all unramified types once we can show that the number of fixed points of g in X_k is bounded as $k \to \infty$. Since this has nothing to do with the choice of a fixed type, we now forget about J and λ (and β, m, etc.) until near the end of this section, and work only with the sets X_k, for *all* $k > 0$.

Note that for each $k' < k$, we get a K-equivariant surjection $X_k \to X_{k'}$. (In fact, these form a projective system

$$X_1 \leftarrow X_2 \leftarrow \cdots$$

of K-sets, and $X = \varprojlim X_k$, but we will not need this fact.) Clearly if g has a fixed point in X_k, then the image of this point in $X_{k'}$ must be a fixed point of g as well.

Lemma. *Assume that $\overline{g}$ is not a scalar. If the characteristic polynomial of $\overline{g}$ is irreducible in $\boldsymbol{k}_F[x]$, then g has at most n fixed points in X_k for all k. Otherwise, g has no fixed points in X_k for all k.*

Proof. Note that X_1 is precisely the subset of $\mathbb{P}^{n-1}(\boldsymbol{k}_E)$ consisting of points whose homogeneous coordinates form a basis of $\boldsymbol{k}_E$ over $\boldsymbol{k}_F$. Furthermore, this set carries the natural action of $\mathrm{GL}_n(\boldsymbol{k}_F) = K/1 + \mathfrak{p}M_n(\mathfrak{o}_F)$, and the action of K on X_1 factors through this quotient. Thus the fixed points of g in X_1 are precisely the fixed points of $\overline{g}$ in X_1, which are just the one-dimensional spaces of eigenvectors of $\overline{g}$ that coincide with points in X_1. Let $p(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_0$ be the characteristic polynomial of g in $\mathfrak{o}_F[x]$, and let $\overline{p}$ be its reduction modulo $\mathfrak{p}$. Suppose that $\overline{p} = p_1p_2$, with p_1 and p_2 relatively prime in $\boldsymbol{k}_F[x]$. Then $\overline{g}$ is similar to a block-diagonal matrix $\left(\begin{smallmatrix} g_1 & \\ & g_2 \end{smallmatrix}\right)$ such that the characteristic polynomial of g_i is p_i for $i = 1, 2$. Clearly such a matrix cannot have a fixed point in X_1.

So if $\overline{g}$ has a fixed point in X_1, $\overline{p}$ must be a power of a single irreducible polynomial. But since n is prime, this means that either $\overline{p}$ is irreducible, or $\overline{p}(x) = (x - \alpha)^n$ for some $\alpha \in \boldsymbol{k}_F^\times$. Assume the latter. Then $\overline{g}$ is conjugate within $\mathrm{GL}_n(\boldsymbol{k}_F)$ to a matrix of the form $\alpha + h$, where h is a nilpotent upper-triangular matrix. Since we are assuming that $\overline{g}$ is not a scalar, $h \neq 0$. Again, it is clear that such a matrix cannot have a fixed point in X_1. This proves the second claim of the lemma.

Now assume that $\overline{p}$ is irreducible. Then by elementary linear algebra, there exists a $\overline{g}$-cyclic vector $\overline{v} \in \boldsymbol{k}_F^n$, i.e., a vector for which

$$\{\overline{v}, \overline{gv}, \ldots, \overline{g}^{n-1}\overline{v}\}$$

is a basis of $\boldsymbol{k}_F^n$. It is not hard to see that any lift v of $\overline{v}$ from $\boldsymbol{k}_F^n$ to $\mathfrak{o}_F^n$ must be g-cyclic, and such a vector yields a basis of F^n consisting of vectors in $\mathfrak{o}_F^n$.

The matrix of g with respect to this basis is the companion matrix

$$C_p = \begin{pmatrix} 0 & & & -a_0 \\ 1 & \ddots & & -a_1 \\ & \ddots & 0 & \vdots \\ & & 1 & -a_{n-1} \end{pmatrix}.$$

Thus g is conjugate within K to C_p, so for the purposes of counting fixed points, we may assume $g = C_p$. Now with g in this simplified form, an easy computation shows that the fixed points of g in X_k correspond precisely to the roots of the polynomial p in $(\mathfrak{o}_E/\mathfrak{p}_E^k)^\times$. If $k = 1$, there are clearly at most n of these. And since $\overline{p}$ is irreducible, they are all distinct, so Hensel's lemma implies that there are at most n such roots in $(\mathfrak{o}_E/\mathfrak{p}_E^k)^\times$ for all $k > 1$ as well. □

The last claim of the theorem, in the unramified case, follows immediately from this lemma. To prove the second claim of the theorem in this case, we return to the context of the beginning of this section of the proof: let (J, λ) be a maximal simple type with all its associated notation, and let $\tau = \mathrm{Ind}_J^K(\lambda)$. Now each of the at most n fixed points of g in X_k corresponds to an $x \in K/J$ such that $x^{-1}gx \in J$. Recalling that $J = \mathfrak{o}_E^\times(1 + \mathfrak{P}^k)$, we see that the reduction modulo $\mathfrak{p}$ of $x^{-1}gx$ will be an element of $(\mathfrak{o}_E/\mathfrak{p}_E)^\times$. Since its minimal polynomial is irreducible of degree n (because it is a conjugate of $\overline{g}$) it must in fact be in $\boldsymbol{k}_E^\times \smallsetminus \boldsymbol{k}_F^\times$. Thus, if m is even, Lemma 2.1 implies that $|\mathrm{Tr}\,\lambda(x^{-1}gx)| \leq 1$. On the other hand, if m is odd, then λ is one-dimensional, so the same is obviously true. Thus either way, the Frobenius formula implies

$$|\mathrm{Tr}\,\tau(g)| \leq n.$$

Finally, we deal with the first claim of the theorem: that the trace of $\tau(g)$ is bounded as τ runs over all unramified types of K. As remarked previously, this will be proved once we show that the number of fixed points of g in X_k is bounded as $k \to \infty$. The only case not covered by the lemma is when $\overline{g}$ is a scalar. For this, we choose $\alpha \in \mathfrak{o}_F^\times$ and $h \in \mathfrak{P}^l = \mathfrak{p}^l M_n(\mathfrak{o}_F)$ such that $g = \alpha + h$, and such that l is maximal with respect to this decomposition. Let ϖ be a uniformizer of F, and let $a = \varpi^{-l}h$. The assumption that l is maximal is equivalent to assuming that $a \in \mathfrak{A} \smallsetminus \mathfrak{P}$ and $\overline{a}$ is not scalar.

Now let $\mathbf{u} = [u_0 : \ldots : u_{n-1}]$ represent a point in X_k for some k. Then this point is fixed by g if and only if

$$\alpha\mathbf{u} + \varpi^l a\mathbf{u} = \gamma\mathbf{u} \pmod{\mathfrak{p}_E^k} \tag{3.3}$$

for some $\gamma \in \mathfrak{o}_E^\times$. Clearly if $k \leq l$, then every point in X_k yields a solution to this, taking $\gamma = \alpha \pmod{\mathfrak{p}_E^k}$. Assume now that $k > l$. Since $a \in \mathfrak{A} \smallsetminus \mathfrak{P}$, this system of equations can have a solution only if we choose $\gamma = \alpha \pmod{\mathfrak{p}_E^l}$, but $\gamma \neq \alpha \pmod{\mathfrak{p}_E^{l+1}}$. Assuming this and letting $\gamma' = \varpi^{-l}(\gamma - \alpha) \in \mathfrak{o}_E^\times$, (3.3) becomes

$$a\mathbf{u} = \gamma'\mathbf{u} \pmod{\mathfrak{p}_E^{k-l}}. \tag{3.4}$$

If $a \notin \mathrm{GL}_n(\mathfrak{o}_F)$, this has no solution, and hence g has no fixed points in X_k for any $k > l$. But if $a \in \mathrm{GL}_n(\mathfrak{o}_F)$, this says that $\mathbf{u}$ represents a fixed point of g in X_k if and only if $\mathbf{u}$ represents a fixed point of a in X_{k-l}. Since $\overline{a}$ is not scalar, the lemma implies that there are at most n such points in X_{k-l}. Thus there are at most $n(\#\boldsymbol{k}_E)^{ln}$ fixed points of g in X_k.

3.3 The ramified case

Now let $\mathfrak{A}$ be the algebra of matrices in $M_n(\mathfrak{o}_F)$ that are upper triangular modulo $\mathfrak{p}$, so that $\mathfrak{A}^\times$ is the Iwahori subgroup of K. As in the previous section of the proof, we temporarily assume all of the notation from the ramified case of section 2.1. In particular, $E = F[\beta]$ is now a totally ramified extension of F of degree n, where $\beta \in M_n(F)$ is minimal over F. Once again, let $k = \left\lfloor \frac{m+1}{2} \right\rfloor$, so that $J = \mathfrak{o}_E^\times(1 + \mathfrak{P}^k)$. Let $\rho = \mathrm{Ind}_J^{\mathfrak{A}^\times}(\lambda)$, and let

$$\tau = \mathrm{Ind}_{\mathfrak{A}^\times}^K(\rho) = \mathrm{Ind}_J^K(\lambda).$$

We will use the same strategy here as in the previous section of the proof, except that we will deal primarily with the induction to $\mathfrak{A}^\times$, which is now a proper subgroup of K.

Let ϖ be a uniformizer of E, and define $X \subset \mathbb{P}^{n-1}(\mathfrak{o}_E)$ to be the set of all points with homogeneous coordinates

$$[u_0 : u_1\varpi : \ldots : u_{n-1}\varpi^{n-1}], \quad u_i \in \mathfrak{o}_E^\times.$$

(Note that this is equivalent to saying the coordinates form an $\mathfrak{o}_F$-basis of $\mathfrak{o}_E$, with strictly increasing E-valuations.) Again it is easy to see that $\mathfrak{A}^\times$ acts transitively on X, and that $\mathfrak{o}_E^\times$ is the stabilizer of some point $x \in X$. A straightforward computation shows that in this case, the normal subgroup $1+\mathfrak{P}^k$ induces the equivalence relation of congruence modulo $\mathfrak{p}_E^k$ on the units

u_i appearing in the coordinates of points in X:

$$\begin{gathered}[u_0 : u_1\varpi : \ldots : u_{n-1}\varpi^{n-1}] \sim [u'_0 : u'_1\varpi : \ldots : u'_{n-1}\varpi^{n-1}] \\ \text{if and only if} \\ u_i = u'_i \pmod{\mathfrak{p}_E^k} \text{ for each } i,\end{gathered}$$

or in other words, congruence modulo $\mathfrak{p}_E^{k+i}$ on the ith coordinate, for each i. If we once again let X_k denote the quotient of X under this equivalence, and let x_k denote the class of x, then $aJ \mapsto a \cdot x_k$ again defines an $\mathfrak{A}^\times$-equivariant bijection

$$\mathfrak{A}^\times/J \to X_k.$$

The same comments apply as before: the actual bijection given above will be different for subgroups J coming from different types, but the action of $\mathfrak{A}^\times$ on the set X_k will be the same regardless; and since the dimension of λ is bounded by a fixed value, we may now forget all about the specific type, and deal only with counting fixed points of g in the sets X_k, for all $k > 0$. Also as before, we have a projective system

$$X_1 \leftarrow X_2 \leftarrow \cdots$$

of $\mathfrak{A}^\times$-sets, with $X = \varprojlim X_k$, and any fixed point of g in X_k must map to a fixed point in $X_{k'}$ for $k' < k$.

We may now dispense easily with the last two claims of the theorem. If $g \in K$ is not K-conjugate to any element of $\mathfrak{A}^\times$, then it clearly cannot be conjugate to any element of J for *any* of the groups J that we are considering. Thus for such an element g, the Frobenius formula implies that $\operatorname{Tr}\tau(g) = 0$ for all ramified types τ of K. On the other hand, if g is conjugate to an element of $\mathfrak{A}^\times$, then for the purpose of computing traces, we may assume $g \in \mathfrak{A}^\times$, and thus $\overline{g} \in \mathrm{GL}_n(\boldsymbol{k}_F)$ is upper-triangular. Clearly such a g can have a fixed point in X_1 only if all of the diagonal entries of $\overline{g}$ (its eigenvalues in $\boldsymbol{k}_F^\times$) are the same, in which case every point of X_1 is a fixed point. Thus, if $\overline{g}$ has at least two distinct eigenvalues, it has no fixed point in X_1, and thus has no fixed point in X_k for all $k > 0$. This proves that $\operatorname{Tr}\rho(g) = 0$ for all types in this case. But since the condition on g here depends only on its conjugacy class in K, applying the Frobenius formula to $\tau = \operatorname{Ind}_{\mathfrak{A}^\times}^K(\rho)$ yields $\operatorname{Tr}\tau(g) = 0$ as well. Note that in this case, the trace bound of n in the second claim of the theorem does not arise at all.

We now deal with the first claim of the theorem. The only remaining possibility for g is that its reduction modulo $\mathfrak{p}$ is upper-triangular with one eigenvalue of multiplicity n. So, just as in the unramified case, we choose $\alpha \in \mathfrak{o}_F^\times$ and $h \in \mathfrak{P}^l$ such that $g = \alpha + h$, and such that l is maximal with respect to this decomposition. In order to describe h more explicitly, let $l = nt + r$ with $0 \le r < n$, and for $0 \le i, j < n$ define

$$\varepsilon_{ij} = \left\lfloor \frac{n-1+r+i-j}{n} \right\rfloor = \begin{cases} 0 & \text{if } r \le j-i \\ 1 & \text{if } r-n \le j-i < r \\ 2 & \text{if } j-i < r-n \end{cases}.$$

Letting h_{ij} denote the i,j coefficient of the matrix h, we may describe h explicitly as follows: (i) $\mathrm{val}_F(h_{ij}) \ge t + \varepsilon_{ij}$ for all i, j, (ii) this inequality is an equality for some i, j satisfying $j - i \equiv r \pmod{n}$, and (iii) if $r = 0$, then the diagonal elements h_{ii} cannot all be the same modulo $\mathfrak{p}^{t+1}$. The first of these requirements is precisely the fact that $h \in \mathfrak{P}^l$; the last two are due to the maximality of our choice of l. Note that $0 \le n\varepsilon_{ij} + j - i - r < n$ for all i, j. So if we let $e_{ij} = n\varepsilon_{ij} + j - i - r$, then e_{ij} is simply the reduction of $j - i - r$ modulo n. From this, we get $\mathrm{val}_E(h_{ij}) \ge l + i - j + e_{ij}$, with equality for some i, j such that $e_{ij} = 0$.

Now let $[u_0 : \varpi u_1 : \ldots : \varpi^{n-1} u_{n-1}]$ represent a point in X_k for some k. Then this point is fixed by g if and only if

$$\alpha \varpi^i u_i + \sum_{j=0}^{n-1} h_{ij} \varpi^j u_j = \gamma \varpi^i u_i \pmod{\mathfrak{p}_E^{k+i}}, \qquad 0 \le i < n, \tag{3.5}$$

for some $\gamma \in \mathfrak{o}_E^\times$. Define a new matrix $a \in M_n(\mathfrak{o}_E)$ by $a_{ij} = \varpi^{j-i-l} h_{ij}$. Then $\mathrm{val}_E(a_{ij}) \ge e_{ij}$, and $a_{ij} \in \mathfrak{o}_E^\times$ for some i, j. The system of equations (3.5) is now equivalent to

$$\varpi^l \sum_{j=0}^{n-1} a_{ij} u_j = (\gamma - \alpha) u_i \pmod{\mathfrak{p}_E^k}, \qquad 0 \le i < n. \tag{3.6}$$

Now it is immediate that if $k \le l$, then any choice of $u_0, \ldots, u_{n-1}$ yields a solution to this system (taking $\gamma = \alpha \pmod{\mathfrak{p}_E^k}$), and hence every point in X_k is a fixed point of g. Assume now that $k > l$. Since $a_{ij} \in \mathfrak{o}_E^\times$ for some i, j satisfying $j - i \equiv r \pmod{n}$, this system has a solution only if we choose $\gamma = \alpha \pmod{\mathfrak{p}_E^l}$ and $\gamma \ne \alpha \pmod{\mathfrak{p}_E^{l+1}}$, and thus only if $a_{ij} \in \mathfrak{o}_E^\times$ for *all* such pairs i, j. If the latter condition is false, then g has no fixed points in X_k and we are finished, so we assume it is true. Let $\mathbf{u}$ denote the column vector in $\mathfrak{o}_E^n$ having $u_0, \ldots, u_{n-1}$ as its components. Also, as in the unramified case, let $\gamma' = \varpi^{-l}(\gamma - \alpha)$. Then (3.6) is equivalent to

$$a\mathbf{u} = \gamma'\mathbf{u} \pmod{\mathfrak{p}_E^{k-l}}. \tag{3.7}$$

We now have two cases to consider. First, if $r = 0$, then $\overline{a} \in \mathrm{GL}_n(\boldsymbol{k}_E)$ is diagonal, but not scalar (since l was chosen to be maximal). Thus, in this case, there can be no fixed points in X_{l+1}, and hence none in X_k for all $k > l$. On the other hand, if $r > 0$, then it is easy to see (since, for example, the matrix a has exactly one unit in each row and column) that there will be exactly one solution to (3.7) for every root of the characteristic polynomial of a in $\mathfrak{o}_E\big/\mathfrak{p}_E^{k-l}$. Since the number of such roots is bounded as $k \to \infty$, the number of solutions to (3.7) is bounded. The fixed points of g in X_k are given by the lifts of these solutions from $\mathfrak{o}_E\big/\mathfrak{p}_E^{k-l}$ to $\mathfrak{o}_E\big/\mathfrak{p}_E^{k}$, and thus are bounded as well. This completes the proof in the ramified case.

Note that in the last case above, the characteristic polynomial of $\overline{a}$ is just $x^n - \overline{\eta}$, where $\eta = \prod a_{ij}$, the product being taken over all i, j such that $j - i \equiv r \pmod{n}$. So we may summarize all of the conditions above as follows: If $r = 0$ or if $\overline{\eta}$ is not an nth power in $\boldsymbol{k}_E^\times$, there will be no fixed points in X_k for all $k > l$. On the other hand, if $r > 0$ and $\overline{\eta}$ is an nth power in $\boldsymbol{k}_E^\times$, then Hensel's lemma again yields (except possibly when n is equal to the residual characteristic) that there are at most n roots in $\mathfrak{o}_E\big/\mathfrak{p}_E^{k}$ for all k. Thus, in this case, there are at most $n(\#\boldsymbol{k}_E)^{ln}$ fixed points of g in X_k for all k, just as in the unramified case. Therefore, in this case, $\mathrm{Tr}(\tau(g))$ is bounded by that number times the maximum dimension of λ (which is $q^{\frac{n^2-n}{2}}$) times $[K : \mathfrak{A}^\times] = \prod_{k=1}^{n-1}(1 + q + \cdots + q^k)$. □

4 The archimedean places

In order to deal with the components of our global types and automorphic representations at the infinite places, we now briefly detour to review some of the representation theory of $\mathrm{U}(n)$ and set up the necessary notation. Let $\mathfrak{g} = \mathfrak{u}(n)$ be the (real) Lie algebra of $\mathrm{U}(n)$, which is the algebra of skew-Hermitian matrices in $M_n(\mathbb{C})$. Let $\mathfrak{g}_\mathbb{R} = i\mathfrak{g}$, the algebra of Hermitian matrices in $M_n(\mathbb{C})$, and let

$$\mathfrak{g}^\mathbb{C} = \mathfrak{g} \otimes_\mathbb{R} \mathbb{C} = \mathfrak{g}_\mathbb{R} \oplus i\mathfrak{g}_\mathbb{R} = \mathfrak{gl}(n, \mathbb{C}).$$

Let T be the maximal torus in $\mathrm{U}(n)$ consisting of diagonal matrices, and let $\mathfrak{h}$ be the corresponding Cartan subalgebra of $\mathfrak{u}(n)$:

$$\mathfrak{h} = \left\{ \begin{pmatrix} ia_1 & & \\ & \ddots & \\ & & ia_n \end{pmatrix} \;\middle|\; a_i \in \mathbb{R} \right\}.$$

As usual, let $\mathfrak{h}_{\mathbb{R}} = i\mathfrak{h}$, and let $\mathfrak{h}^{\mathbb{C}} = \mathfrak{h} \otimes_{\mathbb{R}} \mathbb{C} = \mathfrak{h}_{\mathbb{R}} \oplus i\mathfrak{h}_{\mathbb{R}}$. Finally, let $\mathfrak{h}_{\mathbb{R}}^*$ and $(\mathfrak{h}^{\mathbb{C}})^*$ denote the dual spaces of $\mathfrak{h}_{\mathbb{R}}$ and $\mathfrak{h}^{\mathbb{C}}$, respectively.

To simplify things, we will work relative to a fixed basis: Let e_i be the $n \times n$ matrix with a 1 in the i, i position and zeros elsewhere, so that $\{e_i \mid 1 \le i \le n\}$ is a $\mathbb{C}$-basis of $\mathfrak{h}^{\mathbb{C}}$. We also let e_i^* denote the functionals of the corresponding dual basis, so $e_i^*(e_j) = \delta_{ij}$ for each i, j. Thus any linear functional $\lambda \in (\mathfrak{h}^{\mathbb{C}})^*$ can be written uniquely as $\lambda = \sum_{i=1}^n a_i e_i^*$ $(a_i \in \mathbb{C})$, and such a λ will be analytically integral if and only if $a_i \in \mathbb{Z}$ for all i. In this setting, the set of roots Δ of $\mathrm{U}(n)$ is

$$\Delta = \left\{ \lambda_{ij} = e_i^* - e_j^* \;\middle|\; 1 \le i \ne j \le n \right\}.$$

With respect to our chosen basis of $(\mathfrak{h}^{\mathbb{C}})^*$, the sets of positive and simple roots are, respectively,

$$\begin{aligned} \Delta^+ &= \{\lambda_{ij} \mid 1 \le i < j \le n\} \text{ and} \\ \Pi &= \left\{ \lambda_{i,i+1} = e_i^* - e_{i+1}^* \;\middle|\; 1 \le i < n \right\}. \end{aligned}$$

With these choices, we find that a weight vector $\lambda = \sum a_i e_i^*$ is dominant if and only if $a_i \ge a_j$ for all $i < j$. By the theorem of the highest weight, the irreducible representations of $\mathrm{U}(n)$ are in one-to-one correspondence with the set Λ of dominant, analytically integral functionals on $\mathfrak{h}^{\mathbb{C}}$:

$$\Lambda = \left\{ \lambda = \sum_{i=1}^n a_i e_i^* \;\middle|\; a_i \in \mathbb{Z} \quad \forall i, \text{ and } a_1 \ge \cdots \ge a_n \right\}.$$

For $\lambda \in \Lambda$, we will denote by ξ_λ the corresponding representation of $\mathrm{U}(n)$.

Following standard practice, we define a bilinear form $B_0 : \mathfrak{g} \times \mathfrak{g} \to \mathbb{R}$ by

$$B_0(X, Y) = \operatorname{Tr} XY.$$

Note that for any $\lambda \in \mathfrak{h}_{\mathbb{R}}^*$, there is a unique $H_\lambda \in \mathfrak{h}_{\mathbb{R}}$ such that $\lambda(H) = B_0(H, H_\lambda)$ for all $H \in \mathfrak{h}_{\mathbb{R}}$. We may now define an inner product on $\mathfrak{h}_{\mathbb{R}}^*$ by

$$\langle \lambda_1, \lambda_2 \rangle = B_0(H_{\lambda_1}, H_{\lambda_2}).$$

Let δ be half the sum of the positive roots:

$$\delta = (\tfrac{n-1}{2})e_1^* + (\tfrac{n-3}{2})e_2^* + \cdots + (\tfrac{3-n}{2})e_{n-1}^* + (\tfrac{1-n}{2})e_n^*.$$

The Weyl dimension formula now gives

$$\begin{aligned}\dim(\xi_\lambda) &= \prod_{\alpha\in\Delta^+} \frac{\langle \lambda+\delta, \alpha\rangle}{\langle \delta, \alpha\rangle} \\ &= \prod_{1\le i<j\le n} \frac{a_i - a_j + j - i}{j-i} \\ &= \frac{\prod\limits_{i<j}(a_i - a_j + j - i)}{\prod\limits_{k=1}^{n-1} k!}\end{aligned}$$

for any $\lambda = \sum a_i e_i^* \in \Lambda$. Note that the above expression is a polynomial of degree $\frac{n^2-n}{2}$ in the n variables $a_1, \ldots, a_n$. We will refer to this polynomial as the Weyl polynomial for $\mathrm{U}(n)$.

In order to prove our main global theorem, we will need a bound on the characters of the representations ξ_λ, in analogy with Theorem 3.1. The following proposition is adapted slightly from [**?**, Prop. 1.9], and the proof may be found there.

Proposition 4.1 (Chenevier-Clozel)**.** *Let $g \in \mathrm{U}(n)$, and assume g is not central. There exists a polynomial in n variables $P_g(X_1, \ldots, X_n)$, of degree strictly less than that of the Weyl polynomial, such that for all $\lambda = \sum a_i e_i^* \in \Lambda$,*

$$|\mathrm{Tr}\, \xi_\lambda(g)| \le P_g(a_1, \ldots, a_n).$$

It will be convenient to abuse notation slightly and refer to the Weyl polynomial and the polynomial P_g above as polynomials on $\mathfrak{h}_{\mathbb{R}}^*$, with the understanding that when $\lambda = \sum a_i e_i^*$, $P(\lambda)$ means $P(a_1, \ldots, a_n)$. Note that the degree of such a polynomial is well-defined independently of our choice of basis for $\mathfrak{h}_{\mathbb{R}}^*$.

5 The group G and the global types $\mathcal{T}_G$

We now give a more precise definition of the unitary group G, and the class of global types for G to which our main theorem applies. From here on, F

will denote a totally real number field and E a totally imaginary quadratic extension of F. Let n be prime as before, and let M be a central simple algebra of dimension n^2 over E. Denote by $x \mapsto x^*$ an involution of the second kind of M, i.e., an F-algebra anti-automorphism of M of order 2 whose restriction to E (the center of M) is the non-trivial element of $\mathrm{Gal}(E/F)$. Let G be the unitary group defined (over F) by M and *. Explicitly, this is given by

$$G(R) = \{g \in M \otimes_F R \mid gg^* = 1\} \quad \text{for every } F\text{-algebra } R.$$

In all that follows, we will fix a choice of M and * for which $G(F_v)$ is compact for each infinite place v of F. For each such v, we fix an isomorphism

$$\iota_v : G(F_v) \to \mathrm{U}(n).$$

Let S be the set of places of F which split in E. For each $v \in S$, we will fix an isomorphism

$$\iota_v : G(F_v) \to \mathrm{GL}_n(F_v).$$

For each infinite place v of F, we let $K_v = G(F_v)$, for each $v \in S$, we let $K_v = \iota_v^{-1}(\mathrm{GL}_n(\mathfrak{o}_{F_v}))$, and for each finite place $v \notin S$, we let K_v be any fixed maximal compact subgroup of $G(F_v)$.

Let $\mathbb{A} = \mathbb{A}_\infty \times \mathbb{A}_f$ be the ring of adeles of F, and let $\mathbb{A}_E$ be the ring of adeles of E. Let

$$\begin{aligned} K_\infty &= \prod_{v \mid \infty} K_v = G(\mathbb{A}_\infty), \\ K_f &= \prod_{v \nmid \infty} K_v \subset G(\mathbb{A}_f), \text{ and} \\ K &= K_\infty \times K_f \subset G(\mathbb{A}). \end{aligned}$$

Since G was chosen to be compact at all the infinite places of F, K is actually a maximal compact open subgroup of $G(\mathbb{A})$.

Let Z be the center of G, which is the unitary group of rank 1 defined over F using the extension E/F. This is given explicitly by

$$Z(R) = \{x \in E \otimes_F R \mid xx^* = 1\} \quad \text{for every } F\text{-algebra } R.$$

Note that $Z(F)$ is just $E^1 = \{x \in E^\times \mid N_{E/F}(x) = 1\}$, and similarly $Z(\mathbb{A})$ is just $\mathbb{A}_E^1 = \{x \in \mathbb{A}_E^\times \mid N_{E/F}(x) = 1\}$.

Let Z_0 be the center of K, which is a maximal compact open subgroup of $Z(\mathbb{A})$. Note that the subgroup of rational points of Z_0 is just the group $\mathfrak{o}_E^1$ of units of norm 1 in $\mathfrak{o}_E$, which is simply the finite group μ_E of roots of unity in E.

Let $\mathcal{A}(G(F)\backslash G(\mathbb{A}))$ be the space of automorphic forms on $G(\mathbb{A})$. Since G is compact at infinity, this is simply the space of smooth complex-valued functions on $G(\mathbb{A})$ that are invariant under left translation by elements of $G(F)$, and whose right translates by elements of K span a finite-dimensional space. The group $G(\mathbb{A})$ acts on $\mathcal{A}(G(F)\backslash G(\mathbb{A}))$ by right translation, and the irreducible subquotients of this representation are the automorphic representations of $G(\mathbb{A})$. For an automorphic representation π, we will write $m(\pi)$ for its multiplicity as a composition factor of $\mathcal{A}(G(F)\backslash G(\mathbb{A}))$.

As explained in Section 1.2, we will consider a restricted class of global types for G. To be precise, let $\mathcal{T}_G$ be the set of global types $\tau = \bigotimes_v \tau_v$ of K satisfying the following:

1. For each place v of F, τ_v is an irreducible representation of K_v.

2. For all finite places $v \notin S$ and almost all $v \in S$, $\tau_v = 1$.

3. For all $v \in S$ for which τ_v is not 1-dimensional, $\tau_v = \tau_v' \circ \iota_v$, where τ_v' is the type of a supercuspidal inertial equivalence class for $\mathrm{GL}_n(F_v)$.

For the sake of completeness, we note that the last condition of Definition 1.1, in this context, is

4. If ω_v is the central character of τ_v for each place v, then the character $\omega_\tau = \prod \omega_v$ of Z_0 is trivial on $\mathfrak{o}_E^1$.

For a global type $\tau = \bigotimes_v \tau_v$ of K, it will be convenient to consider its finite and infinite parts separately. We thus write

$$\tau_\infty = \bigotimes_{v\mid\infty} \tau_v \quad \text{and} \quad \tau_f = \bigotimes_{v\nmid\infty} \tau_v,$$

so that τ_∞ is a representation of K_∞, τ_f a representation of K_f, and $\tau = \tau_\infty \otimes \tau_f$.

Now let $\pi = \bigotimes' \pi_v$ be an automorphic representation of $G(\mathbb{A})$. It is clear that the restriction of π to K will contain a global type in $\mathcal{T}_G$ if and only if π_v is either supercuspidal or a twist of an unramified representation at each finite place v of F, and π_v is unramified for each $v \notin S$. Furthermore, under these circumstances, this element of $\mathcal{T}_G$ will be uniquely determined by π, and will occur in π with multiplicity 1.

There is an obvious notion of twisting a global type by a character of K, which is compatible with the twisting of automorphic representations by characters of $G(\mathbb{A})$. Specifically, let θ_v be a character of K_v for each place v, such that $\theta_v = 1$ for almost all finite places v and all $v \notin S$, and such that

$$\theta^n|_{\mathfrak{o}_E^1} = 1,$$

where $\theta = \prod \theta_v$. Then $\theta \otimes \tau$ will be a global type in $\mathcal{T}_G$ as well. We will use the notation $\theta\tau$ for the twist of τ by θ so defined.

Such a character θ of K can always be extended to a unitary character $\chi = \prod \chi_v$ of $G(\mathbb{A})$, for which χ_v will be unramified for almost all $v \in S$ and all finite $v \notin S$, and for which $\chi^n|_{E^1} = 1$. Conversely, given such a character χ of $G(\mathbb{A})$, its restriction θ to K will satisfy all the requirements of the previous paragraph. If π is an automorphic representation of $G(\mathbb{A})$ of type τ, then we can twist π by the character χ to obtain an automorphic representation $\chi\pi$, and clearly it will have type $\theta\tau$. Thus for the purposes of counting automorphic representations of a given type, it will suffice to deal with global types only up to twisting.

6 Proof of main theorem

We are now ready to give the proof of our main result, Theorem 1.2. Since the group G was chosen to be compact at infinity, the proof is a relatively straightforward application of Mackey theory. It can be seen quite easily in the proof that the error term in the theorem comes directly from the trace bounds in Theorem 3.1 and Proposition 4.1, which are combined into a single global lemma below.

Proof of Theorem 1.2. Fix a global type $\tau = \tau_\infty \otimes \tau_f$. We wish to compute $m(\tau)$, the multiplicity of τ in the restriction of $\mathcal{A}(G(F)\backslash G(\mathbb{A}))$ to K. Consider first the isotypic subspace $\mathcal{A}(G(F)\backslash G(\mathbb{A}))^{\tau_\infty}$. This is the space of automorphic forms on $G(\mathbb{A})$ whose right K_∞-translates span a space isomorphic, as a K_∞-representation, to a sum of copies of τ_∞. Note that this subspace is $G(\mathbb{A})$-invariant because of the decomposition $G(\mathbb{A}) = K_\infty \times G(\mathbb{A}_f)$. Clearly $m(\tau)$ is also equal to the multiplicity of τ in $\mathcal{A}(G(F)\backslash G(\mathbb{A}))^{\tau_\infty}$.

For convenience, let W be the space underlying τ_∞, and let $(\tau_\infty^\vee, W^\vee)$ be its contragredient. Let $(\cdot\,,\cdot)$ denote the natural pairing on $W^\vee \times W$. We now consider the space $\mathcal{A}(\,G(F)\backslash G(\mathbb{A})\,, W^\vee)$ of smooth, left-$G(F)$-invariant functions $\phi : G(\mathbb{A}) \to W^\vee$ satisfying

$$\phi(gk) = \tau_\infty^\vee(k^{-1})\phi(g) \quad \text{for all } g \in G(\mathbb{A}) \text{ and } k \in K_\infty.$$

This space affords a natural left action of $G(\mathbb{A}_f)$, by right translation as usual, and the resulting representation can be viewed as the $G(\mathbb{A}_f)$ factor of the space $\mathcal{A}(\,G(F)\backslash G(\mathbb{A})\,)^{\tau_\infty}$. Indeed, given $w \in W$ and $\phi \in \mathcal{A}(\,G(F)\backslash G(\mathbb{A})\,, W^\vee)$, the complex-valued function that maps $g \in G(\mathbb{A})$ to $(\phi(g), w)$ is an automorphic form in $\mathcal{A}(\,G(F)\backslash G(\mathbb{A})\,)^{\tau_\infty}$. This induces a map

$$\tau_\infty \otimes \mathcal{A}(\,G(F)\backslash G(\mathbb{A})\,, W^\vee) \to \mathcal{A}(\,G(F)\backslash G(\mathbb{A})\,)^{\tau_\infty},$$

and it is straightforward to verify that this map is an isomorphism. Therefore, the multiplicity $m(\tau)$ that we seek is equal to the multiplicity of τ_f in the $G(\mathbb{A}_f)$-representation $\mathcal{A}(\,G(F)\backslash G(\mathbb{A})\,, W^\vee)$.

To further simplify things, note that each function in $\mathcal{A}(\,G(F)\backslash G(\mathbb{A})\,, W^\vee)$ is uniquely determined by its restriction to $G(\mathbb{A}_f)$. Thus we consider the space of smooth functions $\phi : G(\mathbb{A}_f) \to W^\vee$ that satisfy

$$\phi(\gamma g) = \tau_\infty^\vee(\gamma)\phi(g) \quad \text{for all } \gamma \in G(F) \text{ and } g \in G(\mathbb{A}_f).$$

Note that this space, with the action of $G(\mathbb{A}_f)$ by right translation, is precisely the induced representation

$$\mathrm{Ind}_{G(F)}^{G(\mathbb{A}_f)}(\tau_\infty^\vee). \tag{6.1}$$

From the comment above, it is clear that restriction of functions yields an isomorphism

$$\mathcal{A}(\,G(F)\backslash G(\mathbb{A})\,, W^\vee) \to \mathrm{Ind}_{G(F)}^{G(\mathbb{A}_f)}(\tau_\infty^\vee).$$

As we are interested in the multiplicity of τ_f in the restriction of this representation from $G(\mathbb{A}_f)$ to K_f, the proof is now reduced to Mackey theory.

Let R be a set of double coset representatives for

$$G(F)\backslash G(\mathbb{A}_f)/K_f\ .$$

Note that

$$G(F)\backslash G(\mathbb{A}_f)/K_f \cong G(F)\backslash G(\mathbb{A})/K\ ,$$

so that R is finite, for example by [**?**, 8.7]. To simplify the notation in what follows, for any $g \in G(\mathbb{A}_f)$ we let $K_{(g)} = G(F)^g \cap K_f$. Note that we may identify $K_{(g)}$ with $G(F)^g \cap K$, considered as a subgroup of $G(\mathbb{A})$, and since $G(F)$ is a discrete subgroup of $G(\mathbb{A})$ and K is compact, $K_{(g)}$ is finite. Applying Mackey's formula to (6.1) now yields

$$\operatorname{Res}_{K_f}^{G(\mathbb{A}_f)} \operatorname{Ind}_{G(F)}^{G(\mathbb{A}_f)}(\tau_\infty^\vee) = \bigoplus_{g\in R} \operatorname{Ind}_{K_{(g)}}^{K_f} \operatorname{Res}_{K_{(g)}}^{G(F)^g}((\tau_\infty^\vee)^g).$$

In what follows, for a compact group H, we will write $\langle \cdot\,, \cdot \rangle_H$ for $\dim \operatorname{Hom}_H(\cdot\,, \cdot)$. Relaxing our notation somewhat, as the restriction functors are implied, we then have

$$\begin{aligned} m(\tau) &= \left\langle \tau_f, \bigoplus_{g\in R} \operatorname{Ind}_{K_{(g)}}^{K_f}((\tau_\infty^\vee)^g) \right\rangle_{K_f} \\ &= \sum_{g\in R} \left\langle \tau_f, \operatorname{Ind}_{K_{(g)}}^{K_f}((\tau_\infty^\vee)^g) \right\rangle_{K_f} \\ &= \sum_{g\in R} \langle \tau_f, (\tau_\infty^\vee)^g \rangle_{K_{(g)}} \\ &= \sum_{g\in R} \langle \tau_\infty^g \otimes \tau_f, 1 \rangle_{K_{(g)}} \\ &= \sum_{g\in R} \frac{1}{\# K_{(g)}} \sum_{x\in K_{(g)}} \operatorname{Tr}(\tau_\infty^g \otimes \tau_f(x)). \end{aligned}$$

Note that $K_{(g)} \cap Z(\mathbb{A}) = G(F) \cap Z_0 = \mathfrak{o}_E^1 = \mu_E$, and by definition a global type τ is assumed to be trivial on this subgroup. Thus in the last sum above, the terms for which x is central all satisfy

$$\operatorname{Tr}(\tau_\infty^g \otimes \tau_f(x)) = \operatorname{Tr}(\tau(x)) = \dim(\tau).$$

Since the group G is compact at infinity, the mass of G with respect to K is simply

$$\mathfrak{m}(G, K) = \sum_{g\in R} \frac{1}{\# K_{(g)}}.$$

Thus we have

$$m(\tau) = \#\mu_E \cdot \mathfrak{m}(G, K) \cdot \dim(\tau) + \sum_{g\in R} \frac{1}{\# K_{(g)}} \sum_{\substack{x\in K_{(g)} \\ x\notin Z(\mathbb{A})}} \operatorname{Tr}(\tau_\infty^g \otimes \tau_f(x)). \quad (6.2)$$

To deal with the second term in this formula, we will apply the following.

Lemma. *Let $g \in G(\mathbb{A}_f)$ and let $x \in K_{(g)} \smallsetminus Z(\mathbb{A})$. Then there exists a constant C_x, and for each infinite place v of F a polynomial $P_{x,v}$ on $\mathfrak{h}_{\mathbb{R}}^*$, such that for all global types $\tau = \tau_\infty \otimes \tau_f \in \mathcal{T}_G$,*

$$|\mathrm{Tr}(\tau_\infty^g \otimes \tau_f(x))| \leq C_x \cdot n^{\#S(\tau)} \cdot \prod_{v|\infty} P_{x,v}(\lambda_v(\tau)).$$

Each of the polynomials $P_{x,v}$ has degree strictly less than that of the Weyl polynomial of $\mathrm{U}(n)$.

Proof. Since x is conjugate to a rational point of G, the eigenvalues of x_v are the same for all v (in the sense that they are roots of the same polynomial over F). Since $K_{(g)}$ is a finite group, x has finite order. But then since $x \notin Z(\mathbb{A})$, x must have at least two distinct eigenvalues. Thus there are at most finitely many places $v \in S$ at which the reduction of x_v modulo $\mathfrak{p}_{F_v}$ has a single eigenvalue of multiplicity n. For each of these places, Theorem 3.1 gives us a constant C_{x_v} such that

$$|\mathrm{Tr}\,\tau_v(x_v)| \leq C_{x_v}$$

for every supercuspidal type τ_v defined on K_v. Let C_x be the product of these constants C_{x_v}. At all other finite places $v \in S$, we have by the same theorem

$$|\mathrm{Tr}\,\tau_v(x_v)| \leq n$$

for all supercuspidal types τ_v defined on K_v. For each infinite place, let $P_{x,v}$ be the polynomial given by Proposition 4.1 applied to $\iota_v(g^{-1}xg) \in \mathrm{U}(n)$. Then for any global type $\tau = \bigotimes_v \tau_v \in \mathcal{T}_G$, since τ_v is 1-dimensional outside of ∞ and $S(\tau)$, we have

$$|\mathrm{Tr}(\tau_\infty^g \otimes \tau_f(x))| = \prod_{v|\infty} \left|\mathrm{Tr}\,\tau_v(g^{-1}xg)\right| \cdot \prod_{v \in S(\tau)} \left|\mathrm{Tr}\,\tau_v(x_v)\right|,$$

and the result follows. □

As there are only finitely many such x to consider, we may sum the constants C_x and the polynomials $P_{x,v}$ in this lemma, and the theorem follows immediately from (6.2). □

The mass of G relative to K appearing in Theorem 1.2 may be computed by a mass formula such as that found in [**?**, 24.4]. We state here a simplified version of that formula, valid when $n > 2$.

Proposition 6.1. *Assume n is odd, let Δ_F (resp. Δ_E) be the absolute discriminant of F (resp. E), and let t be the number of primes of F that are ramified in E. Let χ be the Hecke character of F associated to the field extension E/F by class field theory. Then*

$$\mathfrak{m}(G,K) = 2^{1-t} \cdot \Delta_F^{-\frac{n}{2}} \cdot \left(\frac{\Delta_E}{\Delta_F} (2\pi)^{-[F:\mathbb{Q}]} \right)^{\frac{n^2+n}{2}} \cdot \prod_{k=1}^{n-1} (k!)^{[F:\mathbb{Q}]} \cdot \prod_{k=1}^{n} L(k, \chi^k).$$