ESSENTIAL NUMBER THEORY

2024

vol. 3 no. 1

msp

ESSENTIAL NUMBER THEORY

msp.org/ent

See inside back cover or msp.org/ent for submission instructions.

Essential Number Theory (ISSN 2834-4634 electronic, 2834-4626 printed) at Mathematical Sciences Publishers, 798 Evans Hall #3840, c/o University of California, Berkeley, CA 94720-3840 is published continuously online.

ENT peer review and production are managed by EditFlow® from MSP.

PUBLISHED BY

mathematical sciences publishers

nonprofit scientific publishing

https://msp.org/

Essential Number Theory
Vol. 3, No. 1, 2024
https://doi.org/10.2140/ent.2024.3.1

Invariance of the tame fundamental group under base change between algebraically closed fields

Aaron Landesman

We show that the tame étale fundamental group of a connected normal finite type separated scheme remains invariant upon base change between algebraically closed fields of characteristic $p \geq 0$.

1. Statement of theorem

In a wide range of number theoretic situations, one may want to compare local systems on a variety over one algebraically closed field to local systems on the base change of the variety to a larger algebraically closed field. At least when these local systems are tame, the two notions should be equivalent. Our main result, Theorem 1.1, states that this is indeed true. See Remark 1.6 for some sample uses of this result in number theory.

We now introduce notation to precisely state our main result. Let U be a connected normal finite type separated scheme over an algebraically closed base field k of characteristic p, allowing the possibility $p = 0$. Let $\pi_1(U)$ denote the étale fundamental group of U, where we leave the base point implicit. If $\overline{U}$ is a proper normal scheme containing U as a dense open subscheme, we call $\overline{U}$ a normal compactification of U. If moreover $\overline{U}$ is projective, we call $\overline{U}$ a projective normal compactification of U. Normal compactifications of normal separated finite type schemes always exist, and projective normal compactifications of normal quasiprojective schemes always exist, as described in Remark 1.8.

We next introduce notation to define the numerically tame fundamental group with respect to the above normal compactification $U \to \overline{U}$. We denote this by $\pi_1^{\text{tame}}(U)$, which implicitly depends on the normal compactification $U \subset \overline{U}$. See [Kerz and Schmidt 2010, Appendix, Example 2] for an example demonstrating this dependence on the choice of compactification. Also see Remark 1.9. This numerically tame fundamental group is a quotient of the usual étale fundamental group. Moreover, the prime-to-p étale fundamental group, whose finite quotients correspond to covers of degree relatively prime to p, is a quotient of the tame

MSC2020: 14F35.
Keywords: tameness, fundamental groups, finite étale covers, Abhyankar's lemma, Bertini's theorem.

fundamental group. Here and elsewhere, when $p = 0$, we consider every integer to be relatively prime to p so that the prime-to-p étale fundamental group is the same as the usual étale fundamental group.

First, we introduce the notion of tameness. In order to define tameness, we first recall the definition of the inertia group. Let $E \to U$ be a finite étale Galois G-cover. By convention, we assume Galois covers are connected. Let $s \in \overline{U}$ be a point, let $\overline{E}$ denote the normalization of $\overline{U}$ in the function field of E. Let $t \in \overline{E}$ map to s and define the *decomposition group* of $\overline{E} \to \overline{U}$ at t to be

$$D_{t,\overline{E}/\overline{U}} := \{g \in G : gt = t\}.$$

Then, the *inertia group* of $\overline{E} \to \overline{U}$ at t is $I_{t,\overline{E}/\overline{U}} := \ker(D_{t,\overline{E}/\overline{U}} \to \mathrm{Aut}_s(t))$. Changing our choice of t results in a conjugate inertia group, and so we use $I_{s,\overline{E}/\overline{U}}$ to denote the *inertia group* of $\overline{E} \to \overline{U}$ at s which is the conjugacy class of the subgroup $I_{t,\overline{E}/\overline{U}}$ for any t over s. Note that in the case that the residue fields of s and t agree, the inertia group agrees with the decomposition group. In particular, this automatically holds when the residue field of s is algebraically closed.

We next define tameness. We say $E \to U$ is *tame along* s if the inertia group of $\overline{E} \to \overline{U}$ at s has order prime to p. In the case $E \to U$ is not Galois, we say $E \to U$ is *tame along* s if the Galois closure of $E \to U$ is tame along s. We say $E \to U$ is *tame* if it is tame at every point $s \in \overline{U} - U$.

Finally, we come to the definition of the numerically tame fundamental group. Let $\bar{b} \in U$ denote a geometric point, which we use as a basepoint. For $E \to U$ a finite étale Galois cover, let $\mathrm{Hom}_U(\bar{b}, E)$ denote the set of maps $\bar{b} \to E$ whose composition with $E \to U$ is the given map $\bar{b} \to U$. Following [Kerz and Schmidt 2010, Section 7, page 17] the *numerically tame fundamental group*, $\pi_1^{\mathrm{tame}}(U, \bar{b})$, is by definition the automorphism group of the fiber functor which sends a tame finite étale cover $E \to U$ to $\mathrm{Hom}_U(\bar{b}, E)$. Since every connected finite étale cover is dominated by a Galois finite étale cover, this profinite group is noncanonically in bijection with the profinite set $\lim_{E \to U,\text{ finite étale tame Galois covers}} \mathrm{Hom}_U(\bar{b}, E)$, and the latter is a torsor under the former, whose trivialization can be obtained by choosing a compatible system of basepoints in each $\mathrm{Hom}_U(\bar{b}, E)$. We remind the reader that $\pi_1^{\mathrm{tame}}(U, \bar{b})$ implicitly depends on the choice of normal compactification $U \to \overline{U}$ because the set of finite étale tame Galois covers implicitly depends on the compactification. In what follows, we will omit the basepoint $\bar{b}$ from the notation, and simply write it as $\pi_1^{\mathrm{tame}}(U)$; see [Schmidt 2002; Kerz and Schmidt 2010] for more background on the numerically tame fundamental group. In particular, when k has characteristic 0, $\pi_1(U) \simeq \pi_1^{\mathrm{tame}}(U)$.

If $X \to Y$ and $Z \to Y$ are morphisms, we denote $X \times_Y Z$ by X_Z. In the case $Z = \operatorname{Spec} B$, we also denote $X \times_Y Z$ by X_B.

Our main result is the following theorem.

Theorem 1.1. *Suppose k is an algebraically closed field of characteristic $p \geq 0$ and U is a connected normal separated finite type scheme over k. Let L be any algebraically closed field containing k and $\overline{U}$ any normal compactification of U. Then, the natural map $\pi_1^{\text{tame}}(U_L) \to \pi_1^{\text{tame}}(U)$ is an isomorphism, where tameness for covers of U_L is taken with respect to the normal compactification $U_L \subset (\overline{U})_L$.*

Using the fact that the fundamental group of a scheme is unchanged under inseparable field extensions [Stacks 2005–, Tag 0BQN], we can generalize the above theorem to the case that k and L are only separably closed.

Corollary 1.2. *Suppose k is a separably closed field of characteristic $p \geq 0$ and U is a connected normal separated finite type scheme over k. Let L be any separably closed field containing k and $\overline{U}$ any normal compactification of U. Then, the natural map $\pi_1^{\text{tame}}(U_L) \to \pi_1^{\text{tame}}(U)$ is an isomorphism, where tameness for covers of U_L is taken with respect to the normal compactification $U_L \subset (\overline{U})_L$.*

Remark 1.3. The result Theorem 1.1 for tame fundamental groups described above implies an analogous result for prime-to-p fundamental groups. Namely, let $\pi_1'(U)$ denote the prime-to-p fundamental group, which is the limit of automorphism groups of all Galois finite étale covers of U of degree prime to p. Because prime-to-p covers are all tame, we obtain from Theorem 1.1 that the natural map $\pi_1'(U_L) \to \pi_1'(U)$ is an isomorphism.

Remark 1.4. Theorem 1.1 is surely a folklore theorem. Nevertheless, in its complete form, the author was unable to find it in the literature. The proof written here is primarily a combination of ideas presented to me by Brian Conrad and Jason Starr. In particular, Jason Starr [2016] has written up a separate proof on mathoverflow. The proof in this note is a reorganization of the ideas presented in that post.

Remark 1.5. Many special cases of Theorem 1.1 already exist in the literature. The prime-to-p version of Theorem 1.1 as in Remark 1.3 was previously verified in [Lieblich and Olsson 2010, Corollary A.12] via a proof heavily involving stacks. Separately, this was also shown in [Orgogozo 2003, Corollaire 4.5]. The important special case that U is a curve is also mentioned in [Orgogozo and Vidal 2000, Theorem 6.1], though the proof is omitted there. In characteristic 0, a proof is given in [SGA 1 1971, Exposé XIII, Proposition 4.6] taking $Y = \operatorname{Spec} L$ in the statement there. However, that proof relies on resolution of singularities.

In the case U is proper, this was proven in [Lang and Serre 1957, Théorème 3], [Szamuely 2009, Proposition 5.6.7], [SGA 1 1971, Exposé X, Corollaire 1.8] and also [Stacks 2005–, Tag 0A49].

Remark 1.6. Theorem 1.1 is frequently used in the literature. We provide a few such instances we have come across, but expect that many more examples exist.

In the case U is quasiprojective and k has positive characteristic, the prime-to-p version as in Remark 1.3 is used in [Litt 2021, (4.2.1)] regarding arithmetic representations of fundamental groups.

In the case k has characteristic 0, this result is useful in transferring properties of the fundamental group of a variety over $\mathbb{Q}$ to the corresponding base change to $\mathbb{C}$. For example, this was used in the proof of [Zywina 2010, Lemma 5.2] in order to understand images of Galois representations of abelian varieties. Another sample use is [Landesman 2021, page 701, paragraph 3, proof of Proposition 4.9], where the result was used by the author to estimate average sizes of Selmer groups of elliptic curves over function fields.

As is evident, from the above number theoretic examples, Theorem 1.1 crops up in a variety of situations relevant to number theorists, and so may prove a useful fact in the number theorist's toolkit.

Example 1.7. The tameness hypothesis in the characteristic $p > 0$ case is crucial. If $k \subset L$ are two algebraically closed fields of characteristic $p > 0$, then for U a normal quasiprojective scheme over k, the map $\pi_1(U_L) \to \pi_1(U)$ is not in general an isomorphism. Artin–Schreier covers provide counterexamples in the case $U = \mathbb{A}^1_k$. In more detail, if $\pi_1(\mathbb{A}^1_L) \to \pi_1(\mathbb{A}^1_k)$ were an isomorphism, then the map $H^1(\mathbb{A}^1_k, \mathbb{Z}/(p)) \to H^1(\mathbb{A}^1_L, \mathbb{Z}/(p))$ would also be an isomorphism. The Artin–Schreier exact sequence identifies this with the map

$$k[x]/\{f^p - f : f \in k[x]\} \to L[x]/\{f^p - f : f \in L[x]\},$$

and this map is not surjective because ax^{p-1} for $a \in L - k$ does not lie in the image.

Remark 1.8. Note that the standard definition of the tame fundamental group is more restrictive than our definition in terms of numerical tameness, because the usual definition as in [SGA 1 1971, Expose XIII, 2.1.3] assumes U has a smooth compactification whose boundary is a normal crossings divisor. With this notion from [SGA 1 1971], the tame fundamental group is independent of the choice of compactification.

In contrast, the notion of tame fundamental group we use here makes sense for any normal finite type separated scheme U over k, since we can find a normal compactification of U as follows:

By Nagata compactification, [Stacks 2005–, Tag 0F41] if U is finite type and separated, there exists a quasicompact open immersion $U \to \overline{U}$, where $\overline{U}$ is a proper scheme. One can then replace $\overline{U}$ with its normalization to obtain a proper normal scheme $\overline{U}$, containing U as a dense open.

Moreover, in the case U is quasiprojective, we can also assume $\overline{U}$ is projective by taking any projective scheme $\overline{U}$ containing U as a dense open and then replacing $\overline{U}$ by its normalization.

Remark 1.9. Our notion of the numerically tame fundamental group agrees with the usual notion described in [SGA 1 1971, Expose XIII, 2.1.3] when the compactification of U is smooth with normal crossings boundary by [Schmidt 2002, Proposition 1.14]. This tame fundamental group is not in general independent of the choice of normal compactification; see [Kerz and Schmidt 2010, Appendix, Example 2].

2. Proof of theorem

2.1. ***Idea of proof of Theorem 1.1.*** The proof of Theorem 1.1 is fairly technically involved, but the idea is not too complicated: The key is to verify injectivity of $\pi_1^{\text{tame}}(U_L) \to \pi_1^{\text{tame}}(U)$. As a first step, we reduce from the normal case to the smooth case using that geometrically normal schemes have a dense open smooth subscheme. Then, using Chow's lemma, we reduce to the smooth quasiprojective case. We therefore assume our variety U is smooth and quasiprojective, and prove the theorem by reducing it to the curve case. For this reduction, we fiber U over a variety of one lower dimension, in which case we can apply the curve case to the geometric generic fiber of the fibration.

It remains to deal with the case that U is a quasiprojective smooth curve, which is also the most technically involved part. In this case, we can write U as $\overline{U} - D$, with $\overline{U}$ smooth and projective and D a divisor. To check injectivity, we want to check every finite étale cover of U_L is the base change of some finite étale cover of U. If E is one such cover, we can use spreading out and specialization to obtain an étale cover $U' \to U$ with the same ramification index over each point of D that E has. Then, we construct the cover E' which is the normalization of E in $E \times_{U_L} U'_L$, and verify this is the base change of a cover from k. We do so by applying the projective version of Theorem 1.1, using that E' and U'_L have projective compactifications $\overline{E}'$ and $\overline{U}'_L$ with a finite étale map $\overline{E}' \to \overline{U}'_L$.

We now indicate how we put together the steps described in the above to prove Theorem 1.1. In Section 2.2 (Lemma 2.3), we prove $\pi_1^{\text{tame}}(U_L) \to \pi_1^{\text{tame}}(U)$ is surjective. For injectivity, we first prove the map is injective in the case U is a smooth, connected, and quasiprojective curve in Section 2.4 (Proposition 2.10). We prove in Section 2.14 (Proposition 2.17) that Theorem 1.1 holds for smooth, quasiprojective varieties of all dimensions. We next verify the case that U is smooth, finite type, and separated in Proposition 2.20. Finally, we complete the proof in the case that U is normal, connected, finite type, and separated in Section 2.21.

2.2. ***Surjectivity.*** We first show $\pi_1^{\text{tame}}(U_L) \to \pi_1^{\text{tame}}(U)$ is surjective.

Lemma 2.3. *The map $\pi_1(U_L) \to \pi_1(U)$ is surjective. In particular, $\pi_1^{\text{tame}}(U_L) \to \pi_1^{\text{tame}}(U)$ is surjective.*

Proof. It suffices to verify that the pullback of any connected finite étale cover over U along $U_L \to U$ is connected, see, for example, [Stacks 2005–, Tag 0BN6]. Since L and k are both algebraically closed, the result follows from the fact that connectedness is preserved under base change between algebraically closed fields [EGA IV$_2$ 1965, Proposition 4.5.1]. □

2.4. ***Proof of injectivity in the curve case.*** We next prove injectivity for smooth connected quasiprojective curves U. For this, it suffices to show that any tame Galois finite étale cover E of U_L is the base change of some tame Galois finite étale cover of U. Note that any such cover of U, whose base change is a tame cover E of U_L, is automatically tame, since tameness can be verified after base extension. To prove such an E exists, it suffices to find a connected finite étale cover $F' \to U$ over k so that $F'_L \to U_L$ factors through E.

As a first step, we wish to find a cover U' of U with the same ramification indices as E over points in the normal projective compactification of U.

Notation 2.5. Let $k \to L$ be an inclusion of algebraically closed fields, let U be a smooth curve over k, $\overline{U}$ its regular projective compactification, and $D := \overline{U} - U$. Let $E \to U_L$ be a tame Galois finite étale cover. Let $\overline{E}$ be the normalization of $\overline{U}_L$ inside E.

Lemma 2.6. *With notation as in Notation 2.5, there exists a finite Galois cover $\overline{U}' \to \overline{U}$, étale over U, with the same ramification indices that $\overline{E}$ has over the corresponding points of D_L.*

The idea of this proof is to "spread out and specialize" E. See (2-1) for a diagram.

Proof. To construct $\overline{U}'$, we can find a finitely generated k-subalgebra $A \subset L$ and a finite étale cover $E_A \to U_A$, over A so that $(E_A)_L \simeq E$ and $E_A \to U_A$ is finite étale Galois and tame. Let $\overline{E}_A$ denote the normalization of $\overline{U}_A$ along $E_A \to U_A$. Note that, because of the Galois condition, the ramification index of a point of $\overline{E}_A$ over a point of $\overline{U}_A$ only depends on the image point in $\overline{U}_A$. We may therefore speak of the ramification index over a point of $\overline{U}_A$. Since k is algebraically closed, for any field $K \supset k$, the irreducible components of D_K arise uniquely from the irreducible components of D under scalar extension. We freely use the above observations in what follows.

Let $K(A)$ denote the fraction field of A. Note that the ramification index of $E_{K(A)}$ over each point of $D_{K(A)}$ agrees with that of E over the corresponding point of D_L. Further, we claim that for a general closed point s of Spec A, the ramification index of $s \times_{\operatorname{Spec} A} E_A$ over a point of $s \times_{\operatorname{Spec} A} D_A \simeq D$ agrees with the ramification index of $E_{K(A)}$ over the corresponding generic point of $D_{K(A)}$. To see why this ramification index n is constant over an open set of Spec A, recall that we are assuming the cover $E \to U$ is tame, and so, after possibly shrinking Spec A, we may assume the same of $E_A \to U_A$. By the tameness hypothesis, the ramification index over a point can be identified with one more than the degree of the relative sheaf of differentials at that point; see, for example, [Vakil 2017, page 592]. (The point here is that if the map is locally of the form $t \mapsto us^n$, for t and s uniformizers and u a unit, then the derivative is $dt = d(us^n) = uns^{n-1}ds + s^n du$, which has order precisely $n-1$ if n is not divisible by the characteristic.) So, for $p \in D$ a geometric point, under the identification $p_A \simeq \operatorname{Spec} A$, we see that at any point of $\overline{E}_A \times_{\overline{U}_A} p_A$ over the generic point of $\operatorname{Spec} A$, $\Omega_{\overline{E}_A \times_{\overline{U}_A} p_A / p_A}$ has degree $n-1$. It follows that there is a nonempty open subscheme of $\operatorname{Spec} A$ where $\Omega_{\overline{E}_A \times_{\overline{U}_A} p_A / p_A}$ has degree $n-1$. Hence, the morphism has inertia of order n over some open subscheme of $\operatorname{Spec} A$.

Since k is an algebraically closed field, every closed point of $\operatorname{Spec} A$ has residue field k, so we may choose such a closed point $t\colon \operatorname{Spec} k \to \operatorname{Spec} A$ with the same ramification indices over D as E has over the corresponding points of D_L. Since the locus of geometric points on the base $\operatorname{Spec} A$ where the map $E_A \to U_A$ is a map of connected schemes is constructible [EGA IV$_3$ 1966, Corollaire 9.7.9], we may also assume the fiber of $E_A \to U_A$ over $t : \operatorname{Spec} k \to \operatorname{Spec} A$ is connected. Then, $U' := E_A \times_{\operatorname{Spec} A} \operatorname{Spec} k$ is our desired connected finite étale cover. Finally, we take $\overline{U}'$ to be the normalization of $\overline{U}$ along $U' \to U$. $\square$

Summarizing the situation of Lemma 2.6, we obtain the commutative diagram:

$$\begin{tikzcd}
E \arrow[r] \arrow[d] & E_A \arrow[d] & U' \arrow[l] \arrow[d] \\
U_L \arrow[r] \arrow[d] & U_A \arrow[d] & U \arrow[l] \arrow[d] \\
\operatorname{Spec} L \arrow[r] & \operatorname{Spec} A & \operatorname{Spec} k \arrow[l, "t"']
\end{tikzcd} \tag{2-1}$$

where the four squares are fiber products.

Notation 2.7. Let $\overline{U}' \to \overline{U}$ denote the finite Galois cover of Lemma 2.6. Let $\overline{E}'$ denote the normalization of $\overline{E}$ in $\overline{E} \times_{\overline{U}_L} \overline{U}'_L$ and let $E' := \overline{E}' \times_{\overline{U}'_L} U'_L$, as in the commutative diagram:

$$\begin{tikzcd}
& E' \arrow[rr] \arrow[dl] \arrow[dd] & & \overline{E}' \arrow[dl] \arrow[dd] & \\
E \arrow[rr] \arrow[dd] & & \overline{E} \arrow[dd] & & \\
& U'_L \arrow[rr] \arrow[dl] & & \overline{U}'_L \arrow[dl] & \\
U_L \arrow[rr] & & \overline{U}_L & & D_L \arrow[ll]
\end{tikzcd}$$

Remark 2.8. Observe that the finite map $\overline{U}' \to \overline{U}$ of Notation 2.7 restricts to $U' \to U$ over $U \subset \overline{U}$ as U is normal. By Abhyankar's lemma [Freitag and Kiehl 1988, A I.11] (see also [SGA 1 1971, Expose XIII, 5.2]) we obtain that $\overline{U}'$ is regular, hence smooth, as we are working over an algebraically closed field k.

Although the normalization $\overline{E} \to \overline{U}_L$ of $\overline{U}_L$ in $E \to U_L$ is not necessarily étale, we now show the finite surjection $\overline{E}' \to \overline{U}'_L$ is étale.

Lemma 2.9. *With notation as in Notations 2.5 and 2.7, $\overline{E}' \to \overline{U}'_L$ is étale.*

Proof. Since $E' \to U'_L$ is étale by construction, it is enough to check $\overline{E}' \to \overline{U}'_L$ is étale over all points of $\overline{U}'_L$ lying above a point of D_L. Indeed, this is where we crucially use the assumption that $E \to U$ is tame. Since being étale can be checked in the local ring at each such point, étaleness of $\overline{E}' \to \overline{U}'_L$ follows from a version of Abhyankar's lemma, using that the ramification orders of $\overline{U}'_L \to \overline{U}_L$ and $\overline{E} \to \overline{U}_L$ agree over each point of D_L, by Lemma 2.6. For a precise form of Abhyankar's lemma applicable in this setting; see, for example, [Stacks 2005–, Tag 0EYH]. □

We are now prepared to complete the curve case of Theorem 1.1.

Proposition 2.10. *Theorem 1.1 holds in the case that U is a smooth connected curve.*

Proof. Let U be a smooth connected curve. We use notation from Notation 2.5 and Notation 2.7. By Lemma 2.3, we only need to check injectivity of $\pi_1^{\text{tame}}(U_L) \to \pi_1^{\text{tame}}(U)$. Since $E' \to U_L$ is a finite étale cover of U_L dominating $E \to U_L$, to complete the proof in the case that U is a smooth curve, it suffices to show $E' \to U_L$ is the base change of some tame finite étale cover $F' \to U$ over k. Note here that tameness of $F' \to U$ is automatic once we show it base changes to $E' \to U_L$, as tameness can be verified after base extension. We showed in Lemma 2.9 that $\overline{E}' \to \overline{U}'_L$ is a finite étale cover. Since $\overline{U}'$ is projective and normal, by [Lang and Serre 1957, Théorème 3], we obtain that there is some finite étale cover $\overline{F}' \to \overline{U}'$ over k with $\overline{E}' \simeq (\overline{F}')_L$. (Alternatively, see [Szamuely 2009, Proposition 5.6.7], [SGA 1 1971, Exposé X, Corollaire 1.8], and [Stacks 2005–, Tag 0A49].) We then find $F' := \overline{F}' \times_{\overline{U}'} U'$ is a finite étale cover of U satisfying $(F')_L \simeq E'$, and so this is the desired cover. □

2.11. ***Dominating compactifications.*** In order to complete the reduction from the higher dimensional case to the curve case, we will want to know that Theorem 1.1 holds for one compactification $U \to Y$ whenever it holds for another compactification $U \to X$ with a compatible map $X \to Y$. The next couple lemmas are devoted to verifying this.

Lemma 2.12. *Suppose W is a connected smooth separated finite type scheme over a field k and $\beta : W \subset X$ and $\alpha : W \subset Y$ are two normal compactifications with a map $f : X \to Y$ so that $\alpha = f \circ \beta$. If a finite étale Galois cover $E \to W$ is tame with respect to α, it is also tame with respect to β.*

Proof. Tameness can be checked after field extension, so we will assume k is algebraically closed. Fix a point $s \in Y$ and a preimage $t \in X$ with $f(t) = s$. Let F_Y denote the normalization of Y in the function field of E and let F_X denote the normalization of X in the function field of E. We assume F_Y is tame over s and wish to show F_X is tame over t.

We next claim that there is a map $F_X \to F_Y$. Let F denote the normalization of $F_Y \times_Y X$. It is enough to show the natural map $F \to F_X$ induced by the universal property of normalization is an isomorphism, as we then obtain a map $F_X \simeq F \to F_Y \times_Y X \to F_Y$. Because the normalization map is finite by [Stacks 2005–, Tag 03GR and Tag 035B], both F and F_X are finite over X. Therefore, the map $F \to F_X$ is a birational map which is finite (because it is quasifinite and proper) between normal schemes over k. It follows from a version of Zariski's main theorem that $F \to F_X$ is an isomorphism [Stacks 2005–, Tag 0AB1].

We now conclude the proof. Let v be a point of F_X over t and $u \in F_Y$ be the image of v under the map $F_X \to F_Y$. Since v maps to u, we have an inclusion of decomposition groups $D_{v,F_X/X} \subset D_{u,F_Y/Y}$. Since we are assuming k is algebraically closed, this is identified with an inclusion of inertia groups $I_{v,F_X/X} \subset I_{u,F_Y/Y}$. Hence, up to conjugacy, the inertia group at t is a subgroup of the inertia group at s and so tameness at s implies tameness at t. □

Lemma 2.13. *With the same notation as in Lemma 2.12, if Theorem 1.1 holds with respect to the normal compactification $W \subset X$, Theorem 1.1 also holds with respect to the compactification $W \subset Y$.*

Proof. By Lemma 2.3, it suffices to verify injectivity for the map $\pi_1^{\mathrm{tame}}(W_L) \to \pi_1^{\mathrm{tame}}(W)$ with respect to the compactification $W \subset Y$. Using [Szamuely 2009, Corollary 5.5.8], we can rephrase this as showing that if $k \subset L$ is an extension of algebraically closed fields and $E \to W_L$ is any tame (with respect to $W \to Y$) finite étale Galois cover, then E arises as the base change of a cover $F \to W$ over k. By Lemma 2.12, this cover is also tame with respect to the normal compactification $W \to X$. By assumption Theorem 1.1 holds for the compactification $W \to X$, and so $E \to W_L$ is the base change of a cover $F \to W$ over k, as we wished to show. □

2.14. ***Proof of injectivity in the smooth and quasiprojective case.*** In this section, specifically in Proposition 2.17, we prove Theorem 1.1 in the case that U is a smooth connected quasiprojective variety. To start, we use Bertini's theorem to obtain a fibration away from a codimension 2 subset of U. This fibration will allow us to run an induction on the dimension.

Proposition 2.15. *Let U be a smooth connected quasiprojective variety of dimension $d > 1$. Choose a projective normal compactification $U \subset \overline{U}$. There is a closed subscheme $Z \subset U$ of codimension at least 2 and a projective normal compactification $U - Z \to X$ satisfying the following three properties*:

(1) *The closed subscheme Z lies in the smooth locus of $\overline{U}$.*

(2) *There is a map $X \to \overline{U}$ so that the composition $U - Z \to X \to \overline{U}$ agrees with the composition $U - Z \to U \to \overline{U}$.*

(3) *There is a dominant generically smooth map $\alpha\colon X \to \mathbb{P}^{d-1}_k$ with geometrically irreducible generic fiber.*

Proof. Let $U \subset \overline{U}$ be the given projective normal compactification. Choose an embedding $U \subset \overline{U} \subset \mathbb{P}^n_k$. Replacing $\mathbb{P}^n_k$ by the span of $\overline{U}$ in $\mathbb{P}^n_k$, we may also assume $\overline{U}$ is nondegenerate. Choose a general codimension d plane $H \subset \mathbb{P}^n_k$ such that $H \cap \overline{U}$ is smooth of dimension 0, $H \cap (\overline{U} - U) = \varnothing$, and so that, if $J' \subset \mathbb{P}^n_k$ is a general codimension $d - 1$ plane containing H, we have $J' \cap \overline{U}$ is smooth and geometrically irreducible of dimension 1. This is possible because $\overline{U}$ is normal, hence smooth away from codimension 2, and by Bertini's theorem, as in [Jouanolou 1983, Theoreme 6.10(2) and (3)].

Define $Z := H \cap U = H \cap \overline{U}$ for H as in the previous paragraph. For H general as above, the following three conditions are satisfied: $Z \subset U$ has codimension at least 2, Z does not meet $\overline{U} - U$, and, for a general plane J' containing H, the intersection $J' \cap \overline{U}$ is smooth and geometrically irreducible. The second property verifies condition (1) in the statement because it shows $Z \subset U \subset \overline{U}$ and U is contained in the smooth locus of $\overline{U}$. Take $X \to \overline{U}$ to be the blow up of $\overline{U}$ along $Z \subset \overline{U}$. This verifies condition (2) in the statement.

To conclude, we will show condition (3) in the statement holds. Namely, we will show there is a dominant map $X \to \mathbb{P}^{d-1}_k$ whose generic fiber is smooth and geometrically irreducible. Geometrically, this map is induced by projection of $\overline{U}$ away from the plane H, and sends a point $x \in U - Z$ to $\mathrm{Span}(x, H)$, where we view $\mathrm{Span}(x, H)$ as a point of $\mathbb{P}^{d-1}_k$ parametrizing codimension $d - 1$ planes $J' \subset \mathbb{P}^n_k$ containing H. The above-described map $U - Z \to \mathbb{P}^{d-1}_k$ extends to a map on the blow up $X = \mathrm{Bl}_{\overline{U} \cap H} \overline{U} \to \mathbb{P}^{d-1}_k$, where the fiber over a point $[J'] \in \mathbb{P}^{d-1}_k$ (parametrizing codimension $d - 1$ planes $J' \subset \mathbb{P}^n_k$ containing H) is $J' \cap \overline{U}$. By construction of H so that $J' \cap \overline{U}$ is smooth and geometrically irreducible for a general codimension $d - 1$ plane $J' \subset \mathbb{P}^{d-1}_k$ containing H, the generic fiber of the map $X \to \mathbb{P}^{d-1}_k$ is smooth and geometrically irreducible. $\square$

Assuming we have a fibration as in Proposition 2.15, we next show that the fiber of a tame Galois finite étale cover $E \to U_L$, when restricted to the generic point of $\mathbb{P}^{d-1}_L$, is the base change of a Galois finite étale cover over the generic point of $\mathbb{P}^{d-1}_k$.

Proposition 2.16. *Assume U is a smooth connected k-variety of dimension $d \geq 1$ with a normal projective compactification $U \to \overline{U}$ and a dominant generically smooth map $\alpha\colon \overline{U} \to \mathbb{P}^{d-1}_k$ with geometrically irreducible generic fiber. Let η_k*

denote the generic point of $\mathbb{P}^{d-1}_k$ *and* η_L *denote the geometric generic point of* $\mathbb{P}^{d-1}_L$*. Any given tame finite étale Galois cover* $E \to U_L$ *restricts to a Galois finite étale cover* $E_{\eta_L} \to U_{\eta_L}$ *(with respect to the compactification* $U_{\eta_L} \subset \overline{U}_{\eta_L}$*) which is the base change of some Galois finite étale cover* $F_{\eta_k} \to U_{\eta_k}$*.*

Proof. Let $\bar{\eta}_k$ and $\bar{\eta}_L$ denote compatible algebraic geometric generic points of $\mathbb{P}^{d-1}_k$ and $\mathbb{P}^{d-1}_L$, with corresponding generic points η_k and η_L. By this, we mean that $\bar{\eta}_k$ has residue field which is the algebraic closure of $\kappa(\eta_k)$ and similarly for L. Moreover, they are compatible in the sense that we specify an embedding $\kappa(\bar{\eta}_k) \to \kappa(\bar{\eta}_L)$ restricting to the inclusion $\kappa(\eta_k) \to \kappa(\eta_L)$. Let $E_{\bar{\eta}_L} := E \times_{\mathbb{P}^{d-1}_L} \bar{\eta}_L$, which we note is smooth and of dimension 1. Because $E \to U_L$ is tame with respect to $U_L \to \overline{U}_L$, we obtain that $E_{\bar{\eta}_L} \to U_{\bar{\eta}_L}$ is tame with respect to $U_{\bar{\eta}_L} \to \overline{U}_{\bar{\eta}_L}$. By the curve case of Theorem 1.1, shown in Proposition 2.10, $E_{\bar{\eta}_L}$ arises as the base change of some cover $F_{\bar{\eta}_k} \to U_{\bar{\eta}_k}$. That is, $(F_{\bar{\eta}_k})_{\bar{\eta}_L} \simeq E_{\bar{\eta}_L}$.

To conclude the proof, we only need realize $F_{\bar{\eta}_k} \to U_{\bar{\eta}_k}$ as the base change of a map over η_k so that the above isomorphism $(F_{\bar{\eta}_k})_{\bar{\eta}_L} \simeq E_{\bar{\eta}_L}$ is the base change of an isomorphism over η_L. For K a field, we use K^s to denote its separable closure. We can realize $\bar{\eta}_k \to \eta_k$ as the composition of a purely inseparable morphism $\bar{\eta}_k \to \eta_k^s$ and a separable morphism $\eta_k^s \to \eta_k$ by taking $\eta_k^s := \operatorname{Spec} \kappa(\eta_k)^s$. Since $\bar{\eta}_k \to \eta_k^s$ is a universal homeomorphism, the same is true of $U_{\bar{\eta}_k} \to U_{\eta_k^s}$, and so the map induces an isomorphism of étale fundamental groups $\pi_1(U_{\bar{\eta}_k}) \to \pi_1(U_{\eta_k^s})$ [Stacks 2005–, Tag 0BQN]. It follows that $F_{\bar{\eta}_k} \to U_{\bar{\eta}_k}$ is the base change of a morphism $F_{\eta_k^s} \to U_{\eta_k^s}$ over η_k^s. Moreover, by spreading out, there is a finite Galois extension $\eta'_k \to \eta_k$ so that $F_{\eta_k^s} \to U_{\eta_k^s}$ is the base change of a morphism $F_{\eta'_k} \to U_{\eta'_k}$ over η'_k. We next want to verify this is the base change of a map over η_k, which we will do by producing descent data along the extension $\eta'_k \to \eta_k$.

We next set up notation for descent data. Observe that $\eta_k \simeq \operatorname{Spec} k(x_1, \dots, x_n)$ and $\eta_L \simeq \operatorname{Spec} L(x_1, \dots, x_n)$. Let $M := \Gamma(\eta'_k, \mathcal{O}_{\eta'_k})$ so that $\eta'_k = \operatorname{Spec} M$. It follows that the two maps of schemes $\eta'_k \to \eta_k$ and $\eta_L \to \eta_k$ correspond to the extensions of fields $k(x_1, \dots, x_n) \to M$ and $k(x_1, \dots, x_n) \to L(x_1, \dots, x_n)$. It is a standard fact that these are linearly disjoint, see Lemma A.3. Let $M_L := M \otimes_k L$. Since M and $L(x_1, \dots, x_n)$ are linearly disjoint, base extension defines a bijective map

$$\operatorname{Gal}(M/k(x_1, \dots, x_n)) \simeq \operatorname{Gal}(M_L/L(x_1, \dots, x_n)).$$

We denote the above Galois group by G. As described in [Bosch et al. 1990, Section 6.2, Example B], specifying descent data for $F_{\eta'_k} \to U_{\eta'_k}$ along $\eta'_k \to \eta_k$, is equivalent to specifying an isomorphism $\phi_{F,k,\sigma} : F_{\eta'_k} \to F_{\eta'_k}$ for each $\sigma \in G$, defining an action of G on $F_{\eta'_k}$. (We warn the reader that the action is only defined over η_k and not over η'_k.) Since $U_{\eta'_k}$ is the base change of U_{η_k}, we do have descent data $\phi_{U,k,\sigma} : U_{\eta'_k} \to U_{\eta'_k}$. The descent data $\phi_{F,k,\sigma}$ we wish to produce should live over the descent data for $\phi_{U,k,\sigma}$, in the sense the diagram

$$\begin{array}{ccc} F_{\eta'_k} & \xrightarrow{\phi_{F,k,\sigma}} & F_{\eta'_k} \\ \downarrow & & \downarrow \\ U_{\eta'_k} & \xrightarrow{\phi_{U,k,\sigma}} & U_{\eta'_k} \end{array} \tag{2-2}$$

should commute. Let $\eta'_L := \eta'_k \times_{\eta_k} \eta_L$. Since we do have descent data for $F_{\eta'_L} \to U_{L'}$ along $\eta'_L \to \eta_L$, we have $\phi_{F,L,\sigma}$ and $\phi_{U,L,\sigma}$ so that

$$\begin{array}{ccc} F_{\eta'_L} & \xrightarrow{\phi_{F,L,\sigma}} & F_{\eta'_L} \\ \downarrow & & \downarrow \\ U_{\eta'_L} & \xrightarrow{\phi_{U,L,\sigma}} & U_{\eta'_L} \end{array} \tag{2-3}$$

commutes.

We wish to show that $\phi_{F,L,\sigma}$ is the base change of a unique map $\phi_{F,k,\sigma}$ along $\operatorname{Spec} L \to \operatorname{Spec} k$. Indeed, consider the η_k scheme $\operatorname{Aut}_{\phi_{U,k,\sigma}}(F_{\eta'_k})$ of automorphisms of $F_{\eta'_k}$ over the specified automorphism $\phi_{U,k,\sigma}$ of $U_{\eta'_k}$. Note that $\operatorname{Aut}_{\phi_{U,k,\sigma}}(F_{\eta'_k}) \times_{\eta_k} \eta_L \simeq \operatorname{Aut}_{\phi_{U,L,\sigma}}(F_{\eta'_L})$. Moreover, for $N \in \{k, L\}$, since the automorphisms of $F_{\eta'_N}$ over $\phi_{U,N,\sigma}$ are given by composing any given automorphism over $\phi_{U,N,\sigma}$ with an automorphisms of $F_{\eta'_N}$ over $U_{\eta'_N}$, $\operatorname{Aut}_{\phi_{U,k,\sigma}}(F_{\eta'_k})$ and $\operatorname{Aut}_{\phi_{U,L,\sigma}}(F_{\eta'_L})$ are both G torsors. Since the residue field of each point of $\operatorname{Aut}_{\phi_{U,k,\sigma}}(F_{\eta'_k})$ over η_k is linearly disjoint from the field extension $\kappa(\eta_k) \to \kappa(\eta_L)$ by Lemma A.3, there is a bijection between the points of $\operatorname{Aut}_{\phi_{U,k,\sigma}}(F_{\eta'_k})$ and $\operatorname{Aut}_{\phi_{U,L,\sigma}}(F_{\eta'_L})$. Since the latter is the trivial G torsor, we also obtain $\operatorname{Aut}_{\phi_{U,k,\sigma}}(F_{\eta'_k})$ is the trivial G torsor. In other words there is a unique map $\phi_{F,k,\sigma}$ over $\phi_{U,k,\sigma}$ whose base change to η_L is $\phi_{F,L,\sigma}$. Choosing these $\phi_{F,k,\sigma}$ whose base change is $\phi_{F,L,\sigma}$, we find that the $\phi_{F,k,\sigma}$ define descent data (because the $\phi_{F,L,\sigma}$ do). Hence, $F_{\eta'_k} \to U_{\eta'_k}$ is the base change of a map $F_{\eta_k} \to U_{\eta_k}$, as desired. □

We now complete the proof of Theorem 1.1 in the case U is smooth quasiprojective with a projective normal compactification. Since U is quasiprojective, recall that such a projective normal compactification exists by Remark 1.8.

Proposition 2.17. *Theorem 1.1 holds when $U \to \overline{U}$ is a projective normal compactification and U is smooth and quasiprojective.*

Proof. The case $d = 1$ holds by Proposition 2.10, and $d = 0$ is trivial, so we now assume $d > 1$.

By Proposition 2.15, there is a $Z \subset U \subset \overline{U}$ and a projective normal compactification $U - Z \to X$ satisfying the properties given there. Then, since Z as in Proposition 2.15 has codimension at least 2, $\pi_1^{\text{tame}}(U - Z) \simeq \pi_1^{\text{tame}}(U)$ because the tame fundamental group of a smooth variety is unchanged by removing any set of codimension at least 2, as shown in Lemma A.2. Above, the tameness conditions for both schemes $U - Z$ and U are taken with respect to the projective normal compactification $\overline{U}$.

Observe that Z is in the smooth locus of $\overline{U}$ by Proposition 2.15(1) and $U \to X$ is a normal compactification of U. Using Proposition 2.15(2) to verify the hypotheses of Lemma 2.13, it suffices to prove Theorem 1.1 for the compactification $U - Z \to X$ in place of $U \to \overline{U}$.

For the remainder of the proof, we now rename $U - Z$ as U and X as $\overline{U}$. In particular, by Proposition 2.15(3), we may now assume there is a generically smooth dominant map $\overline{U} \to \mathbb{P}^{d-1}_k$.

With notation as in Proposition 2.16, any tame Galois finite étale cover $E_L \to U_L$ restricts to a cover $E_{\eta_L} \to U_{\eta_L}$ which is the base change of a tame Galois finite étale cover $F_{\eta_k} \to U_{\eta_k}$.

Define F to be the normalization of U in the function field of F_{η_k}. We claim that $F_L \simeq E_L$ as covers of U_L. This will complete the proof, as it implies $F \to U$ is tame finite étale and connected, since the same is true of $F_L \to U_L$.

To see $F_L \simeq E_L$ as covers of U_L, we know E_L is the normalization of U_L in $K(E_L) = K(E_{\eta_L})$. Further, since L/k has a separating transcendence basis (since k is algebraically closed, hence perfect), it follows that F_L is normal and has function field $K(E_L)$. Moreover, the universal property of normalization induces a birational map $F_L \to E$. Since both F_L and E are finite over U_L, the map $F_L \to E$ is finite. It then follows from a version of Zariski's main theorem that $F_L \to E$ is an isomorphism [Stacks 2005–, Tag 0AB1]. □

2.18. ***Proof of injectivity in the smooth case.*** Having verified the smooth quasiprojective case, we next verify the smooth finite type and separated case. The general idea is to use Chow's lemma to reduce to the projective case, but there are a number of technical details. We start by explaining the geometric consequence that Chow's lemma gives us.

Lemma 2.19. *Suppose that U is a smooth separated scheme of finite type over an algebraically closed field k with a normal compactification $\alpha : U \to \overline{U}$. There is a closed subscheme $Z \subset U$ of codimension at least 2 and a normal projective compactification $\beta : U - Z \to X$ with a projective map $f : X \to \overline{U}$ so that $\alpha|_{U-Z} = f \circ \beta$.*

Proof. Using Chow's lemma, we can find a projective scheme X with a birational projective map $f : X \to \overline{U}$; see [Stacks 2005–, Tag 0200 and Tag 0201].

We next construct a subscheme $Z \subset U$ of codimension at least 2 and a birational map $\beta : U - Z \to X$. Since f is birational, there is a dense open $W \subset U$ over which f is an isomorphism, so we obtain a map $g : W \to X$ which is an isomorphism onto its image. Because U is regular in codimension 1 and X is proper, there is a scheme $Z \subset U$ of codimension at least 2 so that $g : W \to X$ extends to a birational map $\beta : U - Z \to X$. Now, restricting f, we get a map $f' : f^{-1}(\alpha(U - Z)) \to U - Z$.

We claim β factors through $f^{-1}(\alpha(U - Z))$ and thus defines a section to f'. Indeed, consider the composition $f \circ \beta : U - Z \to X \to \overline{U}$. This agrees with α over the dense open W, and hence agrees with the given open immersion $U - Z \to U \xrightarrow{\alpha} \overline{U}$ on W. Because $U - Z$ is separated, $f \circ \beta$ must agree with the above open immersion on all of $U - Z$. This implies that β sends $U - Z$ to $f^{-1}(\alpha(U - Z))$.

Let $\beta' : U - Z \to f^{-1}(\alpha(U - Z))$ denote the map whose composition with $f^{-1}(\alpha(U - Z)) \to X$ is β. We will show next that β' is a closed immersion. We have seen above that β' is a section to f'. Therefore, β' is a monomorphism. Moreover since f' is projective, hence proper, β' is also proper, as any section to a proper map is proper via the cancellation theorem [Vakil 2017, 10.1.19] applied to the composition $f' \circ \beta'$. Since β' is a proper monomorphism, it is a closed immersion [Stacks 2005–, Tag 04XV], hence projective.

We now conclude the proof. By the above, the composition $U - Z \to f^{-1}(\alpha(U - Z)) \to X$ is the composition of a closed immersion and an open immersion into a projective scheme. This implies $U - Z$ is quasiprojective, and $U - Z \to X$ is a normal projective compactification, as desired. By construction, $\alpha|_{U-Z} = f \circ \beta$. $\square$

We are now ready to reduce the proof of Theorem 1.1 to the general smooth case over an algebraically closed field, which follows without much difficulty by applying the above lemma.

Proposition 2.20. *Theorem 1.1 holds when U is smooth.*

Proof. Recall that U is now smooth, finite type, and separated over $k = \bar{k}$ but not necessarily quasiprojective. Using Nagata compactification [Stacks 2005–, Tag 0F41] as described in Remark 1.8, we can find a normal compactification $\alpha : U \to \overline{U}$. By Lemma 2.19, there is a closed subscheme $Z \subset U$ of codimension at least 2 and a projective normal compactification $\beta : U - Z \to X$ with a projective map $f : X \to \overline{U}$ so that $\alpha|_{U-Z} = f \circ \beta$.

For $Z \subset U$ of codimension at least 2 as in Lemma 2.19, we have $\pi_1^{\mathrm{tame}}(U) \simeq \pi_1^{\mathrm{tame}}(U - Z)$ by Lemma A.2. Therefore, it is enough to prove the theorem for the compactification $U - Z \to \overline{U}$. By Lemma 2.13, it is enough to prove the theorem for the compactification $U - Z \to X$ in place of $U - Z \to \overline{U}$. Finally, the theorem holds for the projective compactification $U - Z \to X$ by Proposition 2.17. $\square$

2.21. ***Proof of injectivity in the general case.*** We now complete the proof of the theorem for normal connected quasiprojective schemes, using that we have proven it for smooth U.

Proof of Theorem 1.1. By Lemma 2.3, the map $\pi_1^{\mathrm{tame}}(U_L) \to \pi_1^{\mathrm{tame}}(U)$ is surjective. To complete the proof, we wish to show it is injective. To verify the map $\pi_1^{\mathrm{tame}}(U_L) \to \pi_1^{\mathrm{tame}}(U)$ is injective, by [Szamuely 2009, Corollary 5.5.8], it is enough to show that if $E \to U_L$ is any connected finite étale cover, then E is isomorphic to $\widetilde{F}_L$ for $\widetilde{F} \to U$ some connected finite étale cover. To see this, start with some $E \to U_L$. Let $W \subset U$ denote the maximal dense smooth open subscheme of U. Since we have already shown the map $\pi_1^{\mathrm{tame}}(W_L) \to \pi_1^{\mathrm{tame}}(W)$ is an isomorphism in Proposition 2.20, we know that $E \times_{U_L} W_L$ is isomorphic to the base change of some finite étale cover $F \to W$ along $\operatorname{Spec} L \to \operatorname{Spec} k$. Let $\widetilde{F}$ denote the normalization of U in F. Since U is normal, $\widetilde{F} \to U$ is a finite morphism. The setup this far is summarized by the commutative diagrams:

$$\begin{array}{ccc} E \times_{U_L} W_L & \longrightarrow & E \\ \downarrow & & \downarrow \\ W_L & \longrightarrow & U_L \end{array} \qquad \begin{array}{ccc} F & \longrightarrow & \widetilde{F} \\ \downarrow & & \downarrow \\ W & \longrightarrow & U \end{array}$$

To complete the proof, we only need to show $\widetilde{F} \to U$ is tame finite étale and there is an isomorphism $\widetilde{F}_L \simeq E$ over U_L. Indeed, since $\widetilde{F}$ is normal and finite over U, the base change $\widetilde{F}_L$ is also normal and finite over U_L. It follows that $\widetilde{F}_L$ is the normalization of U_L in $F_L \simeq E \times_{U_L} W_L$. But, since E is also the normalization of U_L in $E \times_{U_L} W_L$, we obtain that $E \simeq \widetilde{F}_L$. Since $\widetilde{F}_L \simeq E \to U_L$ is tame finite étale, it follows that $\widetilde{F} \to U$ is also tame finite étale, completing the proof. □

Appendix: Collected lemmas

In this appendix, we collect several lemmas used in the course of the above proof. These are all quite standard, and we only include them for completeness. We include them in this appendix and not in the body so as not to distract from the flow of the proof.

We begin with two standard results on how the tame fundamental group behaves upon passing to open subschemes. These follow from the usual well-known versions for the full étale fundamental group, but we spell out the usual proof for the reader's convenience.

Lemma A.1. *Let Y be a normal quasiprojective connected scheme and $W \subset Y$ be a nonempty open. Then the natural map $\pi_1(W) \to \pi_1(Y)$ is surjective. In particular, $\pi_1^{\mathrm{tame}}(W) \to \pi_1^{\mathrm{tame}}(Y)$ is surjective, where tameness for Y is taken with respect to a projective normal compactification $Y \to \overline{Y}$ and tameness for W is taken with respect to $W \to Y \to \overline{Y}$.*

Proof. Assuming surjectivity of $\pi_1(W) \to \pi_1(Y)$, surjectivity of $\pi_1^{\text{tame}}(W) \to \pi_1^{\text{tame}}(Y)$ follows from commutativity of the square

$$\begin{tikzcd} \pi_1(W) \arrow[r] \arrow[d] & \pi_1(Y) \arrow[d] \\ \pi_1^{\text{tame}}(W) \arrow[r] & \pi_1^{\text{tame}}(Y) \end{tikzcd} \tag{A-1}$$

and the fact that the vertical maps are surjective.

It remains to verify $\pi_1(W) \to \pi_1(Y)$ is surjective. We need to check any connected finite étale cover $E \to Y$ has pullback $E \times_Y W$ which is also connected. First, we claim E is normal. Indeed, since normality is equivalent to being R1 and S2, E is normal because the properties of being R1 and S2 are preserved under étale morphisms. Therefore, E is normal and connected, hence integral. Then, $E \times_Y W$ is a nonempty open subscheme of the integral scheme E, hence connected. □

For a proof of the next lemma in the case of fundamental groups, instead of tame fundamental groups; see [Szamuely 2009, Corollary 5.2.14].

Lemma A.2. *Let U be a connected smooth k-scheme and $V \subset U$ a closed subscheme of codimension at least* 2. *Then the natural map $\pi_1^{\text{tame}}(U - V) \to \pi_1^{\text{tame}}(U)$ is an isomorphism, where tameness for U is taken with respect to a projective normal compactification $U \to \overline{U}$, and tameness for $U - V$ is taken with respect to $U - V \to U \to \overline{U}$.*

Proof. The map is surjective by Lemma A.1, so it suffices to verify injectivity. For this, we have to show that any tame finite étale cover $E \to U - V$ extends uniquely to a tame finite étale cover E' of U. If $E \to U - V$ is tame, it follows from the definition of tameness and our compatible choices of compactifications that any extension will automatically also be tame. Hence, it suffices to show there is a unique extension. Uniqueness is immediate because E' is necessarily normal, and hence must be the normalization of U in E. So it suffices to check that the normalization E' of U in E is a finite étale cover of U, restricting to E over $U - V$. That E' restricts to E over $U - V$ is clear and $E' \to U$ is finite by finiteness of normalization. Finally, $E' \to U$ is étale by Zariski–Nagata purity as in [SGA 1 1971, Exposé X, Théorème 3.1] because it is étale over all codimension 1 points and U is smooth. □

Finally, we record a field-theory result on linear disjointness of certain extensions.

Lemma A.3. *Suppose $k \to L$ are algebraically closed fields. Let $k(x_1, \ldots, x_n) \to F$ by any finite separable extension. Then $k(x_1, \ldots, x_n) \to F$ and $k(x_1, \ldots, x_n) \to L(x_1, \ldots, x_n)$ are linearly disjoint extensions.*

Proof. We want to show the only finite separable extension of $k(x_1, \ldots, x_n)$ in $L(x_1, \ldots, x_n)$ is $k(x_1, \ldots, x_n)$. To this end, let F be some finite separable extension of $k(x_1, \ldots, x_n)$ in $L(x_1, \ldots, x_n)$. So, to see F is equal to $k(x_1, \ldots, x_n)$, it suffices to show $F \otimes_{k(x_1,\ldots,x_n)} F$ is a domain. We have a containment

$$F \otimes_{k(x_1,\ldots,x_n)} F \subset L(x_1, \ldots, x_n) \otimes_{k(x_1,\ldots,x_n)} L(x_1, \ldots, x_n),$$

so it suffices to show

$$L(x_1, \ldots, x_n) \otimes_{k(x_1,\ldots,x_n)} L(x_1, \ldots, x_n)$$

is a domain. Indeed, this is a localization of

$$L[x_1, \ldots, x_n] \otimes_{k[x_1,\ldots,x_n]} L[x_1, \ldots, x_n] \simeq (L \otimes_k L)[x_1, \ldots, x_n],$$

so it suffices to show $L \otimes_k L$ is a domain. This then holds because L is a domain, and a domain over an algebraically closed field is still a domain upon base change to any larger algebraically closed field, i.e., the property of being geometrically integral is preserved under base change between algebraically closed fields. $\square$

Acknowledgements

I would like to thank Brian Conrad, Jason Starr, and Sean Cotner for key ideas in the proof. I also thank several anonymous referees for numerous incredibly thorough readings, many extremely helpful comments, and multitudes of thoughtful suggestions. Additionally, I thank Sean Cotner for a detailed reading, and thorough comments. I thank Peter Haine, Daniel Litt, Martin Olsson, Tamás Szamuely and the referees for helpful comments. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-1656518.

References

[Bosch et al. 1990] S. Bosch, W. Lütkebohmert, and M. Raynaud, *Néron models*, Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)] **21**, Springer, 1990. MR Zbl

[EGA IV$_2$ 1965] A. Grothendieck, "Éléments de géométrie algébrique, IV: Étude locale des schémas et des morphismes de schémas, II", *Inst. Hautes Études Sci. Publ. Math.* **24** (1965), 5–231. MR Zbl

[EGA IV$_3$ 1966] A. Grothendieck, "Éléments de géométrie algébrique, IV: Étude locale des schémas et des morphismes de schémas, III", *Inst. Hautes Études Sci. Publ. Math.* **28** (1966), 5–255. MR Zbl

[Freitag and Kiehl 1988] E. Freitag and R. Kiehl, *Étale cohomology and the Weil conjecture*, Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)] **13**, Springer, 1988. MR Zbl

[Jouanolou 1983] J.-P. Jouanolou, *Théorèmes de Bertini et applications*, Progress in Mathematics **42**, Birkhäuser, Boston, 1983. MR Zbl

[Kerz and Schmidt 2010] M. Kerz and A. Schmidt, "On different notions of tameness in arithmetic geometry", *Math. Ann.* **346**:3 (2010), 641–668. MR Zbl

[Landesman 2021] A. Landesman, "The geometric average size of Selmer groups over function fields", *Algebra Number Theory* **15**:3 (2021), 673–709. MR Zbl

[Lang and Serre 1957] S. Lang and J.-P. Serre, "Sur les revêtements non ramifiés des variétés algébriques", *Amer. J. Math.* **79** (1957), 319–330. MR Zbl

[Lieblich and Olsson 2010] M. Lieblich and M. Olsson, "Generators and relations for the étale fundamental group", *Pure Appl. Math. Q.* **6**:1 (2010), 209–243. MR Zbl

[Litt 2021] D. Litt, "Arithmetic representations of fundamental groups, II: Finiteness", *Duke Math. J.* **170**:8 (2021), 1851–1897. MR Zbl

[Orgogozo 2003] F. Orgogozo, "Altérations et groupe fondamental premier à p", *Bull. Soc. Math. France* **131**:1 (2003), 123–147. MR Zbl

[Orgogozo and Vidal 2000] F. Orgogozo and I. Vidal, "Le théorème de spécialisation du groupe fondamental", pp. 169–184 in *Courbes semi-stables et groupe fondamental en géométrie algébrique* (*Luminy*, 1998), edited by J.-B. Bost et al., Progr. Math. **187**, Birkhäuser, Basel, 2000. MR Zbl

[Schmidt 2002] A. Schmidt, "Tame coverings of arithmetic schemes", *Math. Ann.* **322**:1 (2002), 1–18. MR Zbl

[SGA 1 1971] A. Grothendieck, *Revêtements étales et groupe fondamental* (Séminaire de Géométrie Algébrique du Bois Marie 1960–1961), Lecture Notes in Math. **224**, Springer, 1971. MR Zbl

[Stacks 2005–] "The Stacks project", 2005–, available at http://stacks.math.columbia.edu.

[Starr 2016] J. Starr, "Is the map on etale fundamental groups of a quasi-projective variety, upon base change between algebraically closed fields, an isomorphism?", MathOverflow thread, 2016, available at http://mathoverflow.net/q/257762.

[Szamuely 2009] T. Szamuely, *Galois groups and fundamental groups*, Cambridge Studies in Advanced Mathematics **117**, Cambridge University Press, 2009. MR Zbl

[Vakil 2017] R. Vakil, "The rising see: Foundations of algebraic geomerty", preprint, 2017, available at http://math.stanford.edu/~vakil/216blog/FOAGnov1817public.pdf.

[Zywina 2010] D. Zywina, "Hilbert's irreducibility theorem and the larger sieve", preprint, 2010. arXiv 1011.6465v1

Received 15 Jul 2022. Revised 19 Dec 2023.

AARON LANDESMAN:

landesman@math.harvard.edu
Department of Mathematics FAS, Harvard University, Cambridge, MA, United States

msp

msp

Essential Number Theory
Vol. 3, No. 1, 2024
https://doi.org/10.2140/ent.2024.3.19

Weak transfer from classical groups to general linear groups

Sug Woo Shin

Following Arthur, we present a trace formula argument proving that discrete automorphic representations on (possibly non-quasisplit) classical groups weakly transfer to general linear groups in the sense that the transfer is compatible with Satake parameters and infinitesimal characters. This result is conditional on the weighted fundamental lemma but no more. We explain how the weak transfer leads to the existence of automorphic Galois representations valued in the C-groups, as formulated by Buzzard and Gee, when the automorphic representations are C-algebraic and satisfy suitable regularity conditions.

1. Introduction

Classical groups are the isometry groups of symmetric, symplectic or (skew-) Hermitian forms. They play vital roles in many areas of mathematics. In number theory they are prominent in the theory of automorphic forms and the Langlands program. One of the key questions is how to transfer automorphic representations on classical groups to general linear groups as predicted by the Langlands functoriality conjecture. There are two main approaches: the converse theorem and the trace formula.

The converse theorem was successfully employed to transfer cuspidal generic automorphic representations on quasisplit classical groups over number fields by Cogdell, Kim, Krishnamurthy, Piatetski-Shapiro, Shahidi, and others; see [Cogdell et al. 2011]. Lomeli [2009] proved the analogous result for split classical groups over global function fields. There is a prospect, arising from the work by Cai,

MSC2020: 11F70, 11F72, 11R39.

Keywords: Langlands functoriality, Langlands correspondence, trace formula, automorphic representations, Galois representations.

Friedberg, Ginzburg and Kaplan [Cai et al. 2019], that the converse theorem method may extend to all classical groups without any genericity condition.

It is perhaps fair to say that the trace formula method requires more groundwork to get started, notably the stabilization of the trace formula and the fundamental lemma as well as their twisted analogues. Since the tools are still developing over global function fields, see [Labesse and Lemaire 2021], we will concentrate on the *number field case* throughout the paper. When it works, the trace formula leads to extra information beyond the existence of transfer to general linear groups, such as parametrization of local and global packets of representations characterized by endoscopic character identities and the Arthur multiplicity formula. This has been carried out for

- quasisplit symplectic and special orthogonal groups by Arthur [2013];
- quasisplit unitary groups by Mok [2015];
- non-quasisplit unitary groups by Kaletha, Minguez, Shin and White [2014], under temperedness and pure-inner-twist hypotheses;
- non-quasisplit odd special orthogonal groups by Ishimoto [2023], under a temperedness hypothesis;
- certain non-quasisplit symplectic and special orthogonal groups under a cohomological hypothesis at infinity by Taïbi [2019].

It is worth mentioning that Clozel and Labesse (see [Labesse 2011]) proved unconditional results on the transfer of cohomological automorphic representations on unitary groups to those on general linear groups (without full endoscopic classifications for them). However the results in the bulleted list are conditional on the proof of the weighted fundamental lemma and some results to be proven. (By "some results", we mean the projected papers in [Arthur 2013], which the author cites as [A25], [A26] and [A27], as well as their analogues for unitary groups, which are also missing at the time of writing this article.) The weighted fundamental lemma is known for split groups by Chaudouard and Laumon [2010; 2012] but it is also needed for nonsplit groups. We also need the "nonstandard weighted fundamental lemma" formulated by Waldspurger [2009] in the stabilization of the twisted trace formula. See the paragraph above Theorem 1.1.2 for further remarks.

Apart from the conditionality mentioned above, the trace formula is believed to yield complete results for all non-quasisplit classical groups as outlined in [Arthur 2013, Chapter 9]. This is a central problem to work out in its own right. It is also pivotal for arithmetic applications involving Shimura varieties since non-quasisplit groups appear naturally in that context. A full solution of the problem would take years to complete.

The first goal of this paper is to explain that Arthur's argument [2013, Chapter 3] is already enough to establish the existence of a weak transfer for *all* classical groups. He states the results for quasisplit symplectic and special orthogonal groups but the argument works generally. Indeed, Arthur himself [2013, Proposition 9.5.2] made this observation; our intention is merely to bring this part of his work to the broader audience.

Here a weak transfer means a transfer of automorphic representations between two reductive groups related via a morphism of their L-groups, such that the Satake parameters at finite places and the infinitesimal characters at infinite places are transported via the L-morphism; see Section 1.1 below. Our argument is relatively simple as long as the stabilization of the twisted (and untwisted) trace formula is accepted. In particular we do not need [A25], [A26] and [A27] from [Arthur 2013], or their analogues mentioned above (nor the main theorems of [Arthur 2013; Mok 2015]). Rather, the weak transfer at hand is conditional only on the weighted fundamental lemma for nonsplit groups and the nonstandard weighted fundamental lemma.

Our approach to the weak transfer is close to Taïbi's [2022], see Remark A.5 therein. The difference is that his argument and theorem are optimized for the intended application. As such, he accepts the main results of [Arthur 2013] and makes a regularity hypothesis to deal with non-quasisplit symplectic and special orthogonal groups. By contrast, we keep a minimal hypothesis as mentioned above and also treat the case of unitary groups in a uniform manner.

As an application and our second goal, we verify Buzzard and Gee's conjecture on the existence of automorphic Galois representations, which amounts to one direction of the global Langlands correspondence, for classical groups. Besides the weak transfer, a crucial ingredient comes from what is known in the construction of automorphic Galois representations for general linear groups. Once this is taken for granted, it is a series of elementary exercises to deduce Buzzard and Gee's conjecture for classical groups (modulo some technical hypotheses discussed below). While we do not claim originality, it may be of interest to see all classical groups treated side by side in the language of C-groups. Previous works usually considered these groups separately; e.g., see [Kret and Shin 2020, Section 6; 2023, Section 2] and the references at the start of Section 3.4 below.

Now we describe the two main goals more precisely in Sections 1.1 and 1.2 below. They correspond to Sections 2 and 3 in the main body of the paper.

1.1. ***Weak transfer.*** Let G and $\tilde{G}$ be connected reductive groups over a number field F, and $\tilde{\xi} : {}^L G \to {}^L \tilde{G}$ be a morphism of L-groups (either the Galois or Weil form, see [Arthur 2005, Section 26]). Assume that $\tilde{G}$ is quasisplit over F. Let S be a finite set of places of F including all infinite places such that G, $\tilde{G}$, and $\tilde{\xi}$ are unramified over F_v for all places $v \notin S$. (For $\tilde{\xi}$, this means that $\tilde{\xi}$ is inflated from an L-morphism with respect to the Galois or Weil group for an extension unramified at v.) At each $v \notin S$, the map $\tilde{\xi}$ induces a map $\tilde{\xi}_*$ from irreducible unramified representations of $G(F_v)$ to those of $\tilde{G}(F_v)$ (on the level of isomorphism classes) by Satake transform, which amounts to the unramified local Langlands correspondence for each of G and $\tilde{G}$.

A weak form of the Langlands functoriality conjecture is the following, see [Langlands 1970, Questions 3 and 5] and the commentary in [Arthur 2021, Section 4] for instance.

Conjecture 1.1.1. *Let $\tilde{\xi} : {}^L G \to {}^L \tilde{G}$ be a morphism of L-groups. For each automorphic representation π of $G(\mathbb{A}_F)$, there exists an automorphic representation Π of $\tilde{G}(\mathbb{A}_F)$ such that, for every $v \notin S$ where π is unramified, Π_v is unramified and isomorphic to $\tilde{\xi}_*(\pi_v)$. Moreover the infinitesimal characters of archimedean components of Π are determined by those of π via $\tilde{\xi}$.*

If Π as above exists, we say that Π is a *weak transfer* (a.k.a. a weak functorial lift) of π. It is said to be weak because the conjecture does not address what happens at the places in S nor what the set of all Π as above looks like. A stronger conjecture can be best formulated in terms of local Arthur packets at *all* places as well as global Arthur packets, as accomplished in the endoscopic classification for classical groups mentioned above. By focusing on the weak version, we bypass the subtlety of Arthur packets at the expense of losing precision.

We are particularly interested in Conjecture 1.1.1 where π appears in the discrete spectrum of the space of L^2-automorphic forms on $G(\mathbb{A}_F)$. Although the beyond endoscopy program was proposed by Langlands to attack this conjecture, the general case is still completely out of reach. Good news is that substantial progress has been made in the (twisted) endoscopic case, namely when $\tilde{\xi}$ realizes G as a (twisted) endoscopic group for $\tilde{G}$. A prominent example is Langlands and Arthur and Clozel's base change [1989] for general linear groups, where $G = \mathrm{GL}_n$ and $\tilde{G} = \mathrm{Res}_{F'/F}\,\mathrm{GL}_n$ (Weil restriction of scalars) for a finite solvable extension F'/F. See [Cogdell 2003, Section 4] for more on the base change and other examples.

This paper is concerned with a weak transfer for classical groups. In this case G is a classical group and $\tilde{G}$ is (the restriction of scalars of) a general linear group; the latter is denoted $\tilde{G}^0(N)$ in the main text. We are divided into Cases S and U:

Case S: G is a special orthogonal or a symplectic group, $\tilde{\xi}$ is the standard embedding.

Case U: G is a unitary group and $\tilde{\xi}$ is the base change embedding (up to a twist).

In these two cases the quasisplit inner form G^* of G may be thought of as a twisted endoscopic group for $\tilde{G}$; see Sections 2.1 and 2.2 for more details. Henceforth we make the following hypothesis as our method crucially relies on the stabilization of the (possibly twisted) trace formula by Arthur and Moeglin and Waldspurger:

(H1) The weighted fundamental lemma (WFL) is true for nonsplit groups. Moreover its nonstandard version is true.

It is worth elaborating on the hypothesis. The stabilization of the twisted trace formula [Mœglin and Waldspurger 2016a; 2016b] requires the twisted weighted fundamental lemma [Mœglin and Waldspurger 2016a, II.4.4], which is reduced by the main result of [Waldspurger 2009] to the WFL for Lie algebras and the nonstandard WFL. The latter two, precisely formulated in Sections 3.6 and 3.7 of [Waldspurger 2009], assert certain identities of weighted orbital integrals on the Lie algebras of two reductive groups which are related by endoscopic data or nonstandard endoscopic data, respectively. As mentioned above, the WFL for Lie algebras remains to be verified for nonsplit groups. The nonstandard WFL is open at this time.

With that said, hypothesis (H1) can be black-boxed since we only need the outcome of the stabilization, namely (2.4.4) and (2.4.6) below. Let us state our first main theorem.

Theorem 1.1.2. *Assuming (H1), Conjecture 1.1.1 is true for Cases S and U above.*

Here is the idea of proof in the essential case when $G = G^*$, i.e., when G is quasisplit; see the proof of Theorem 2.5.1 for complete details. By induction, we may assume that the theorem is known for all classical groups of smaller rank, or finite products thereof. Let π be as in Conjecture 1.1.1. Let c^S and ζ denote the family of Satake parameters of π away from S and the infinitesimal character of π at ∞, respectively. The L-morphism $\tilde{\xi}$ transfers c^S and ζ to a family of Satake parameters $\tilde{c}^S$ and an infinitesimal character $\tilde{\zeta}$ for $\tilde{G}$. We assume that $(\tilde{\zeta}, \tilde{c}^S)$ does not appear in the automorphic spectrum for $\tilde{G}$. The goal is to derive a contradiction.

The main input is the stabilized trace formula relating G and $\tilde{G}$, where the subscript $\tilde{\zeta}, \tilde{c}^S$ indicates the $(\tilde{\zeta}, \tilde{c}^S)$-isotypic part of each trace formula (reviewed in Section 2.4 following [Arthur 2013, Chapter 3]; we recommend [Arthur 2005] for a detailed introduction to the trace formula)

$$I^{\tilde{G}}_{\mathrm{disc},\tilde{\zeta},\tilde{c}^S}(f) = \sum_{G^{\tilde{\mathfrak{e}}}} \iota(\tilde{\mathfrak{e}}) S^{\tilde{\mathfrak{e}}}_{\mathrm{disc},\tilde{\zeta},\tilde{c}^S}(f^{\tilde{\mathfrak{e}}}), \tag{1.1.1}$$

where:

- $I^{\tilde{G}}_{\mathrm{disc}}$ is an invariant distribution on $\tilde{G}(\mathbb{A}_F)$, which is the discrete part of the invariant trace formula for the twisted group $\tilde{G}$.
- $G^{\tilde{\mathfrak{e}}}$ stands for the twisted endoscopic group in a twisted elliptic endoscopic datum $\tilde{\mathfrak{e}}$ for $\tilde{G}$ (up to isomorphism); this includes $G^{\tilde{\mathfrak{e}}} = G$.
- $\iota(\tilde{\mathfrak{e}}) \in \mathbb{Q}$ is a positive constant.
- $S^{\tilde{\mathfrak{e}}}_{\mathrm{disc}}$ is a stable distribution on $G^{\tilde{\mathfrak{e}}}(\mathbb{A}_F)$, which is the discrete part of the stable trace formula for the twisted endoscopic group of $\tilde{\mathfrak{e}}$.

- f is a decomposable test function on $\tilde{G}(\mathbb{A}_F)$ whose components away from S belong the unramified Hecke algebras.
- $f^{\tilde{\mathfrak{e}}}$ is a function on $G^{\tilde{\mathfrak{e}}}(\mathbb{A}_F)$ which is a transfer of f.

Although $S^{\tilde{\mathfrak{e}}}_{\mathrm{disc},\tilde{\zeta},\tilde{c}^S}$ is very complicated in general, the induction hypothesis can be used to show that $S^{\tilde{\mathfrak{e}}}_{\mathrm{disc},\tilde{\zeta},\tilde{c}^S}$ is equal to the trace on the $(\tilde{\zeta},\tilde{c}^S)$-isotypic part of the L^2-discrete spectrum of $\tilde{G}^{\tilde{\mathfrak{e}}}$. The point is that the "error terms" (the difference between the two quantities in the preceding sentence) all come from classical groups of smaller rank, which have to do with automorphic representations of general linear groups by induction, whereas $(\tilde{\zeta},\tilde{c}^S)$ is unrelated to such representations by hypothesis. In particular, for $\tilde{\mathfrak{e}}$ such that $G^{\tilde{\mathfrak{e}}}=G$, the stable distribution $S^{\tilde{\mathfrak{e}}}_{\mathrm{disc},\tilde{\zeta},\tilde{c}^S}$ is *not* the zero distribution since π appears in the sum. (Recall that $(\tilde{\zeta},\tilde{c}^S)$ is the image of (ζ,c^S) via $\tilde{\xi}$.)

The left-hand side of (1.1.1) is trivially zero by the assumption that $(\tilde{\zeta},\tilde{c}^S)$ does not appear in the automorphic spectrum of $\tilde{G}$. Hence our preceding observation about $S^{\tilde{\mathfrak{e}}}_{\mathrm{disc},\tilde{\zeta},\tilde{c}^S}$ tells us that a certain *nonnegative* combination of traces of irreducible representations on different groups on the right-hand side vanishes. We crucially invoke Arthur's vanishing result [2013, Section 3.5], exactly designed for these circumstances and relying on the nonnegativity of coefficients, to show that the right-hand side is term-by-term trivial, i.e., every nonnegative coefficient is zero. This is a contradiction since $S^{\tilde{\mathfrak{e}}}_{\mathrm{disc},\tilde{\zeta},\tilde{c}^S}$ was seen to be nontrivial.

1.2. ***Automorphic Galois representations.*** For the moment we go back to a general connected reductive group G over a number field F. An automorphic representation π of $G(\mathbb{A}_F)$ is called L-algebraic (resp. C-algebraic) if the infinitesimal character of π at ∞ is algebraic (resp. algebraic after shifting by the half sum of positive roots), see Definition 3.1.1 below. By ${}^CG$ we denote the C-group of G introduced by Buzzard and Gee [2014], which is a certain semiproduct of ${}^LG$ with $\mathbb{G}_m$; see Section 3.1 below. It can also be thought of as the L-group of a central $\mathbb{G}_m$-extension of G.

Fix a prime ℓ. Let S denote the finite set of places of F containing all ℓ-adic and infinite places as well as the finite places v such that either G or π is ramified at v. When $v\notin S$, write

$$\phi_{\pi_v}=\phi^L_{\pi_v}:W_{F_v}\to {}^LG$$

for the unramified Langlands parameter for π_v, with coefficient in $\mathbb{C}$. We also define a C-normalized parameter

$$\phi^C_{\pi_v}:W_{F_v}\to {}^CG$$

by modifying ϕ_{π_v}; see below Lemma 3.1.5 for more details. In this paper, a *Galois representation* $\Gamma_F\to {}^LG(\overline{\mathbb{Q}}_\ell)$ or $\Gamma_F\to {}^CG(\overline{\mathbb{Q}}_\ell)$ always means a continuous semisimple representation which is unramified at all but finitely places and whose restriction to the local Galois group at each place above ℓ is de Rham. When the de Rham condition is satisfied, the Galois representations can be assigned Hodge–Tate cocharacters (Section 3.1).

Buzzard and Gee [2014] formulated the following, see Conjectures 3.1.2 and 3.1.8 below, generalizing from the case of general linear groups in Clozel's work [1990].

Conjecture 1.2.1. *Let $? \in \{L, C\}$, ℓ a prime, and $\iota : \mathbb{C} \cong \overline{\mathbb{Q}}_\ell$ an isomorphism. For each $?$-algebraic discrete automorphic representation π of $G(\mathbb{A}_F)$, there exists a Galois representation*

$$r = r_{\ell,\iota}(\pi) : \Gamma_F \to {}^{?}G(\overline{\mathbb{Q}}_\ell)$$

such that:

(i) $r|^{\mathrm{ss}}_{W_{F_v}} \cong \iota\phi^{?}_{\pi_v}$ *at finite places $v \notin S$.*

(ii) *The Hodge–Tate cocharacters of r are explicitly determined by the infinitesimal characters of π at ∞.*

Our interest lies in the conjecture when G is a classical group. We will concentrate on the C-algebraic case for two reasons. Firstly, it is more directly related to the geometric Satake equivalence (that is, part (i) of the conjecture is compatible with geometric Satake in the C-algebraic case, see [Zhu 2020b]) and the cohomology of Shimura varieties (e.g., as observed in [Johansson 2013]). Secondly, the C-algebraic case is more general as illustrated by the example of an even unitary group (i.e., of even rank) over a totally real field relative to a CM quadratic extension. Such a group does not possess any L-algebraic automorphic representations whose archimedean components belong to discrete series whereas there are many C-algebraic ones. (In fact, one can go from the C-algebraic case to the L-algebraic case and vice versa after pulling back via a central $\mathbb{G}_m$-extension of G, see [Buzzard and Gee 2014, Section 5], but we do not discuss it further.) With that said, it is worth mentioning that C-algebraicity and L-algebraicity coincide for symplectic, even special orthogonal, and odd unitary groups.

From now, assume that F is a totally real field. In Case U, assume that G is a unitary group with respect to a CM quadratic extension E over F, and write $c \in \mathrm{Gal}(E/F)$ for the nontrivial element. In Case S, set $E := F$ and $c := 1$ (trivial automorphism of F).

We fix π as in Theorem 1.1.2, so the theorem provides us with an automorphic representation Π of $\mathrm{GL}_N(\mathbb{A}_E)$ for a suitable N. Without loss of generality we assume that Π is an isobaric sum of cuspidal automorphic representations of smaller general linear groups: $\Pi = \boxplus_{i=1}^{r} \Pi_i$. (In fact we show that Π can be chosen as such when proving the theorem.) By the strong multiplicity one theorem, such a Π is unique up to isomorphism. (Hence $\Pi_1, \dots, \Pi_r$ are unique up to isomorphism and permutation.) For each i, we write $\Pi_i^\star$ for the contragredient of $\Pi_i \circ c$, where c naturally acts on $\mathrm{GL}_N(\mathbb{A}_E)$. Consider the following hypotheses:

(H2) The infinitesimal character of Π is regular at infinity, see Definition 3.2.1 below.

(H3) Each Π_i is (conjugate) self-dual, i.e., $\Pi_i^\star \cong \Pi_i$ for every i.

Condition (H2) is equivalent to regularity of the infinitesimal character of π at infinity unless G^* is an even special orthogonal group (Lemma 3.2.2). Hypothesis (H3) is implied by a full endoscopic classification theorem, which is a conditional theorem for classical groups as already discussed. Our second main theorem is the following (Theorem 3.2.7).

Theorem 1.2.2. *Assume (H1), (H2), and (H3). Then the C-algebraic version of Conjecture 1.2.1 holds true in Cases S and U above, except that* (*i*) *is true only up to outer automorphism in the even orthogonal case. If we assume only (H1) and (H2) then we have the existence of the Galois representation as in the conjecture satisfying* (*i*) *but possibly not* (*ii*).

Let us outline the steps of the proof:

(Step 1) Prove Conjecture 1.2.1 for cuspidal regular automorphic representations Π_0 of GL_N over totally real or CM fields (see Proposition 3.1.11 below for the precise version).

(Step 2) Combine Step 1 with Theorem 1.1.2 to construct a GL_N-valued Galois representation $R(\pi)$ corresponding to given π on a classical group.

(Step 3) Factor the Galois representation $R(\pi)$ through the L or C-group of G. In Case U, this entails extending the Galois representation along the quadratic extension E/F.

Step 1 follows by combining the work of many authors as recalled in the proof of Proposition 3.1.11, if Π_0 is moreover (conjugate) self-dual up to a character. Without hypothesis (H3), we need to appeal to more recent work by Harris, Lan, Taylor and Thorne [Harris et al. 2016] and Scholze [2015]. In this case we lose control of the Hodge–Tate cocharacter. (See the last paragraph in the proof of Proposition 3.1.11.) This is why part (ii) of Conjecture 1.1.1 is not verified when (H3) is not assumed. Other than this, the argument is the same whether (H3) is assumed or not.

In Step 2 we start from a weak transfer $\pi \mapsto \Pi = \boxplus_{i=1}^r \Pi_i$ and apply Step 1 to construct Galois representations R_i from Π_i. The desired Galois representation is essentially $\bigoplus_{i=1}^r R_i$ but this is not literally true. We need to keep a careful track of L and C-normalizations.

In Step 3 the main input is Bellaïche and Chenevier's [2011] result on the sign of Galois representations. Thanks to this, the argument is relatively simple in Case S. More work is needed in Case U, but knowing the sign again allows us to factor the extended Galois representation through the C-group.

Remark 1.2.3. When F is a global function field of characteristic $p > 0$, if $\ell \neq p$ then Conjecture 1.2.1 can be stated in terms of the L-group of G, without imposing condition (ii) or algebraicity. (Every automorphic representation is considered algebraic.) Then Conjecture 1.2.1 is true for every G and every cuspidal π by V. Lafforgue [2018].

1.3. ***Complements.*** We comment on the prospect of removing hypotheses (H1), (H2), and (H3). The author is cautiously optimistic that the removal of (H1) would be attainable within the next few years. It may be possible to weaken the regularity condition (H2) in Theorem 1.2.2 to weak regularity of Π at infinity in the sense of [Fakhruddin and Pilloni 2019, Section 9.1]; the weak regularity (and oddness) of Π is always satisfied if π has regular infinitesimal character at infinity, even when G is an even special orthogonal group. A crucial input is [Boxer and Pilloni 2021, Theorem 6.11.2], which relaxes the regularity assumption on π in Proposition 3.1.11 to weak regularity. The proof of Proposition 3.2.4, except the assertions on signs, goes through with the weakening of (H2) as long as both (H1) and (H3) are assumed. The only missing ingredient is the analogue of the main results of [Bellaïche and Chenevier 2011] when Π is weakly regular (and odd) but not regular. To remove (H3), the main problem is to compute the Hodge–Tate weights of the automorphic Galois representations in [Harris et al. 2016; Scholze 2015] as mentioned above. Partial results are available in [A'Campo 2024, Theorem 1.0.6; Hevesi 2023, Theorem 1.1].

There are other ways to strengthen Theorems 1.1.2 and 1.2.2. Theorem 1.1.2 is going to be eventually superseded by a full endoscopic classification; the point of our theorem lies in the simplicity and uniformity of the argument. Theorem 1.2.2 can be upgraded by listing more properties satisfied by the Galois representation r. For instance, we can ask for a description of the image of complex conjugation at real places of F, see Remark 3.2.8. Another question is to prove local-global compatibility at *all* finite places v, namely that the Weil–Deligne representation associated with r at v corresponds to the v-component of the automorphic representation via the local Langlands correspondence. This is known in the setting of Proposition 3.1.11 for GL_N. (If π is not conjugate self-dual up to a character then the compatibility is known away from places above ℓ.) From this, our existing arguments should justify the local-global compatibility for G at all finite places (avoiding places above ℓ if (H3) is not assumed), at least if G is quasisplit. In fact, such a reasoning already appears in the proof of [Kret and Shin 2023, Theorem 2.4 (i), (iv)] and [Kret and Shin 2020, Theorem 6.4(SO-i)] in some special cases. If G is not quasisplit then the same should work once the local Langlands correspondence for G becomes available in a way that is compatible with the local Langlands for its quasisplit inner form.

Finally one can try to characterize those Galois representations which correspond to automorphic representations in Conjecture 1.2.1. In fact it is fruitful to view the Galois representations as global L-parameters and extend the Galois representations to some sort of global A-parameters as in [Johansson and Thorne 2020, Section 4]. Then a natural problem is to formulate local and global A-packet classifications for algebraic automorphic representations by means of such Galois-theoretic A-parameters. We hope to address this elsewhere.

1.4. ***Notation and conventions.*** Let k be a perfect field. Denote by $\bar{k}$ an algebraic closure of k. Write $\Gamma_{k'/k} := \mathrm{Gal}(k'/k)$ for any Galois extension k'/k and put $\Gamma_k := \Gamma_{\bar{k}/k}$. When T is a torus over k, write $X^*(T) := \mathrm{Hom}_{\bar{k}}(T, \mathbb{G}_m)$ and $X_*(T) := \mathrm{Hom}_{\bar{k}}(\mathbb{G}_m, T)$. Put $X^*(T)_R := X^*(T) \otimes_{\mathbb{Z}} R$ for $\mathbb{Z}$-algebras R, which is an $R[\Gamma_k]$-module. Define $X_*(T)_R$ likewise. Let $\widehat{T}$ denote the dual torus of T over $\mathbb{C}$ equipped an action of Γ_k.

From now on, let F be a number field. Write $\mathbb{A}_F$ for the ring of adèles and $\mathbb{A}_F^S$ for the ring of adèles away from S, where S is a finite set of places of F. For each place v of F, write W_{F_v} for the local Weil group. We fix the embeddings $\iota_v : \overline{F} \hookrightarrow \overline{F}_v$ at each v, which induce the injections $\Gamma_{F_v} \hookrightarrow \Gamma_F$. If v is a complex place, then there are two $\mathbb{R}$-isomorphisms $\iota_1, \iota_2 : \overline{F}_v \cong \mathbb{C}$. For each complex embedding $\tau : F \hookrightarrow \mathbb{C}$ inducing the place v, we write $\iota_\tau : \overline{F} \hookrightarrow \mathbb{C}$ for either $\iota_1\iota_v$ or $\iota_2\iota_v$, whichever induces τ via the inclusion $F \subset \overline{F}$. If τ is a real embedding inducing v then set $\iota_\tau := \iota_v$. Thus we have $\iota_\tau : \overline{F} \hookrightarrow \mathbb{C}$ extending every embedding $\tau : F \hookrightarrow \mathbb{C}$.

Let F_0 be a subfield of F (allowing $F_0 = F$), and S a finite set of places of F_0 containing all infinite places. Then $\Gamma_{F,S}$ denotes the Galois group $\mathrm{Gal}(F_S/F)$, where $F_S \subset \overline{F}$ is the maximal extension of F which is unramified at every place of F which lies above some place of F_0 in S.

Let G^* be a connected quasisplit reductive group over F, with an F-pinning $(B^*, T^*, \{X_\alpha^*\})$. Let $\widehat{G}^*$ denote the Langlands dual group over $\mathbb{C}$ equipped with a Γ_F-action on $\widehat{G}^*$ (called an L-action), a Γ_F-pinning $(\widehat{B}^*, \widehat{T}^*, \{\widehat{X}_{\alpha^\vee}^*\})$, and a Γ_F-equivariant bijection between the based root datum of $\widehat{G}^*$ and the dual based root datum of G^*. This allows us to define the Galois form of the L-group

$$^LG^* := \widehat{G}^* \rtimes \Gamma_F.$$

It is also convenient to use $\Gamma_{F'/F}$ in place of Γ_F, where F' is a finite extension of F over which G^* splits. Only in Section 2 we will occasionally consider the Weil form of the L-group, with the Weil group of F in place of Γ_F. We will often fix an isomorphism $\iota : \mathbb{C} \cong \overline{\mathbb{Q}}_\ell$ and also view $\widehat{G}^*$ and $^LG^*$ over $\overline{\mathbb{Q}}_\ell$. Write $S_{\mathrm{bad}}(G^*)$ for the set of places v of F which are either infinite or such that $G^*_{F_v}$ is ramified.

At $v \notin S_{\text{bad}}(G^*)$, the pinning determines a hyperspecial subgroup $K_v^* \subset G^*(F_v)$. Unramified representations of $G^*(F_v)$ at $v \notin S_{\text{bad}}(G^*)$ are always meant to be relative to this K_v^*.

Let G be a connected reductive group over F with an isomorphism $\mathfrak{i} : G^*_{\overline{F}} \simeq G_{\overline{F}}$ such that $\mathfrak{i}^{-1}\sigma(\mathfrak{i})$ is an inner automorphism of $G^*_{\overline{F}}$ for every $\sigma \in \Gamma_F$. Such a pair $(G, \mathfrak{i})$ is called an *inner twist* of G^* over F, and classified up to isomorphism by the Galois cohomology valued in the adjoint group $H^1(F, G^{*,\text{ad}})$, whose image in $H^1(F_v, G^{*,\text{ad}})$ is trivial for v not contained a finite set of places S. Then $H^1(F, G^{*,\text{ad}}(\mathbb{A}_F^S)) = \bigoplus_{v \notin S} H^1(F_v, G^{*,\text{ad}})$ is trivial, so $\mathfrak{i}$ is defined over $\mathbb{A}_F^S$ after conjugation by an element of $G^{*,\text{ad}}(\mathbb{A}_F^S)$. Thereby we obtain an isomorphism $G^*(\mathbb{A}_F^S) \cong G(\mathbb{A}_F^S)$, canonical up to $G^*(\mathbb{A}_F^S)$-conjugacy. Put $S_{\text{bad}}(G) := S_{\text{bad}}(G^*) \cup S$. At each $v \notin S_{\text{bad}}(G)$, we transport hyperspecial subgroups K_v^* to $K_v \subset G(F_v)$ via the isomorphism and use them for the notion of unramified representations. We transfer the F-pinning for G^* to a pinning for G via $\mathfrak{i}$ so that the based root data for G^* and G are Γ_F-equivariantly identified. Thereby we may and will identify the L-group ${}^LG$ with ${}^LG^*$, and transfer $(\widehat{B}^*, \widehat{T}^*, \{\widehat{X}^*_{\alpha^\vee}\})$ for $\widehat{G}^*$ to $(\widehat{B}, \widehat{T}, \{\widehat{X}_{\alpha^\vee}\})$ for $\widehat{G}$.

For a place v of G, we often write G_v to mean $G \times_F F_v$. Write $F_\infty := F \otimes_{\mathbb{Q}} \mathbb{R} = \prod_{v \mid \infty} F_v$, and $G_\infty := (\operatorname{Res}_{F/\mathbb{Q}} G) \times_{\mathbb{Q}} \mathbb{R} = \prod_{v \mid \infty} G_v$. We fix a maximal compact subgroup $K_\infty = \prod_{v \mid \infty} K_v \subset G_\infty(\mathbb{R}) = \prod_{v \mid \infty} G(F_v)$.

By $\mathcal{H}(G)$ we denote the space of smooth compactly supported functions on $G(\mathbb{A}_F)$ which are bi-K-finite under some compact subgroup $K = \prod_v K_v \subset G(\mathbb{A}_F)$, where K_v is the fixed hyperspecial subgroup (resp. maximal compact subgroup) at all but finitely many v (resp. all infinite places v). Let $\mathcal{H}(G_\infty)$ denote the space of smooth compactly supported bi-K_∞-finite functions on $G_\infty(\mathbb{R})$. Let S be a finite set of finite places of F containing $S_{\text{bad}}(G)$. At $v \notin S$, let $\mathcal{H}_{\text{ur}}(G_v)$ denote the unramified Hecke algebra of bi-K_v-invariant functions on $G(F_v)$. Take $\mathcal{H}^S_{\text{ur}}(G)$ to be the unramified Hecke algebra of compactly supported bi-K^S-invariant functions on $G(\mathbb{A}_F^S)$, where $K^S = \prod_{v \notin S} K_v$ is the product of fixed hyperspecial subgroups. The analogous definition of $\mathcal{H}(G)$, possibly with decorations, makes sense when G is a nontrivial coset in a twisted group, e.g., $G = G(N)$ as in Section 2.2 below.

Write A_G for the maximal $\mathbb{Q}$-split torus in the center of $\operatorname{Res}_{F/\mathbb{Q}} G$. (We have $A_G = \{1\}$ for the classical groups to be considered.) Put

$$[G] := G(F)\backslash G(\mathbb{A}_F)/A_G(\mathbb{R})^0.$$

Let $L^2_{\text{disc}}([G])$ denote the discrete part of the L^2-space of functions on $[G]$, viewed as a $G(\mathbb{A}_F)$-module by right translation. Every irreducible $G(\mathbb{A}_F)$-subrepresentation is referred to as a discrete automorphic representation. Denote by $L^2_{\text{disc}}([G])^{S-\text{ur}}$ the subspace generated by discrete automorphic representations which are unramified away from S. Write $\mathcal{C}^S(G)$ for the set in which each member is a family of semisimple $\widehat{G}$-conjugacy classes $c_v \subset {}^L G_v$ over finite places $v \notin S$ such that c_v maps to the geometric Frobenius element under the projection from ${}^L G_v$ to the unramified Galois group over F_v. By the Satake isomorphism, each c_v corresponds to a $\mathbb{C}$-algebra morphisms $\mathcal{H}_{\mathrm{ur}}(G_v) \to \mathbb{C}$ at $v \notin S$. Thereby $\mathcal{C}^S(G)$ is identified with the set of $\mathbb{C}$-algebra morphisms $\mathcal{H}^S_{\mathrm{ur}}(G) \to \mathbb{C}$.

Write $G_{\infty,\mathbb{C}} := (\mathrm{Res}_{F/\mathbb{Q}}\, G) \times_{\mathbb{Q}} \mathbb{C} = \prod_{\tau : F \hookrightarrow \mathbb{C}} G_\tau$, where $G_\tau := G \times_{F,\tau} \mathbb{C}$. Let $T_{\infty,\mathbb{C}} = \prod_\tau T_\tau$ be a maximal torus in $G_{\infty,\mathbb{C}}$. The Lie algebra of $T_{\infty,\mathbb{C}}$ is denoted by $\mathfrak{t}_{\infty,\mathbb{C}}$. Write $\Omega_\infty = \prod_\tau \Omega_\tau$ for the Weyl group of $T_{\infty,\mathbb{C}}$ in $G_{\infty,\mathbb{C}}$. We often write Ω for Ω_τ for simplicity.

We use $\mathfrak{Z}(G_\infty)$ to denote the center of the universal enveloping algebra of the Lie algebra of $G_{\infty,\mathbb{C}}$. By the Harish-Chandra isomorphism, we may identify $\mathfrak{Z}(G_\infty) = \mathbb{C}[\mathfrak{t}_{\infty,\mathbb{C}}]^\Omega$. Write $\mathcal{C}_\infty(G)$ for the set of $\mathbb{C}$-algebra morphisms $\mathfrak{Z}(G_\infty) \to \mathbb{C}$, or equivalently

$$\mathcal{C}_\infty(G) = \mathfrak{t}^*_{\infty,\mathbb{C}}/\Omega = X^*(T_\infty)_{\mathbb{C}}/\Omega_\infty = X_*(\widehat{T}_\infty)_{\mathbb{C}}/\Omega_\infty = \prod_\tau X_*(\widehat{T}_\tau)_{\mathbb{C}}/\Omega. \quad (1.4.1)$$

Let $\pi = \otimes'_v \pi_v$ be an irreducible admissible representation of $G(\mathbb{A}_F)$ such that π is unramified outside S. At each $v \notin S$, each π_v corresponds to a semisimple $\widehat{G}$-conjugacy class $c(\pi_v) \subset {}^L G_v$ known as the Satake parameter of π_v, and vice versa. By assigning to π the infinitesimal character at ∞ and the Satake parameters away from S, we obtain a map

$$\pi \mapsto (\zeta_{\pi_\infty}, (c(\pi_v))_{v \notin S}) \in \mathcal{C}_\infty(G) \times \mathcal{C}^S(G).$$

According to the decomposition (1.4.1), we write

$$\zeta_{\pi_\infty} = (\zeta_{\pi_\infty,\tau})_{\tau : F \hookrightarrow \mathbb{C}}.$$

For π as above, we have an unramified L-parameter $\phi_{\pi_v} : W_{F_v} \to {}^L G_v$ at each $v \notin S$ and an archimedean L-parameter $\phi_{\pi_v} : W_{F_v} \to {}^L G_v$ at $v \mid \infty$. The relation to the above map is as follows. For $v \notin S$, ϕ_{π_v} sends lifts of the geometric Frobenius element into $c(\pi_v)$. For $v \mid \infty$ and each $\tau : F \hookrightarrow \mathbb{C}$ inducing v, if we identify $\overline{F}_v = \mathbb{C}$ via τ thus $W_{\overline{F}_v} = \mathbb{C}^\times \subset W_{F_v}$, then $\phi_{\pi_v}|_{\mathbb{C}^\times}$ is $\widehat{G}$-conjugate to a map of the form

$$z \in \mathbb{C}^\times \mapsto \lambda(z)\lambda'(\bar{z}) \in \widehat{T}_\tau \subset \widehat{G}_\tau = \widehat{G}_v$$

such that $\lambda = \zeta_{\pi_\infty,\tau}$.

When v is a place of F, we denote by $|\cdot|_v$ the usual norm character on $F_v^\times$ or W_{F_v} valued in positive real numbers, satisfying the product formula. Our normalization at finite places v is that a uniformizer in $F_v^\times$ and a lift of the geometric Frobenius in W_{F_v} both map to the inverse of the residue field cardinality. By $\det_N : \mathrm{GL}_N \to \mathbb{G}_m$ we mean the determinant map, and $|\det_N|_v : \mathrm{GL}_N(F_v) \to \mathbb{R}_{>0}$ the map $x \mapsto |\det_N(x)|_v$. We often omit N and v and simply write $|\cdot|$, det, and $|\det|$.

Given a finite dimensional representation r (typically of a local Weil group), r^{ss} stands for its semisimplification. By an (ℓ-adic) *Galois representation* of Γ_F, where F is a number field, we mean a continuous semisimple representation of Γ_F on a finite-dimensional $\overline{\mathbb{Q}}_\ell$-vector space which is unramified at almost all places of F and de Rham at ℓ. More generally, when G is as above, an ${}^LG$ or ${}^CG$-valued *Galois representation* is a continuous representation

$$\Gamma_F \to {}^LG(\overline{\mathbb{Q}}_\ell) \quad \text{or} \quad R : \Gamma_F \to {}^CG(\overline{\mathbb{Q}}_\ell)$$

which:

- Is unramified at almost all places of F.
- Commutes with the projections from Γ_F and the L or C-groups onto the Galois group $\Gamma_{F'/F}$, where F'/F is a Galois extension with respect to which ${}^LG$ or ${}^CG$ is formed.
- $i \circ R$ is semisimple and de Rham at ℓ for i a faithful algebraic representation (see [Borel 1979, Section 2.6]) of the L-group or C-group.

For G over F as above, write $\mathcal{E}_{\mathrm{ell}}(G)$ for a set of representatives for isomorphism classes of (standard) elliptic endoscopic data $(H, \mathcal{H}, s, \xi)$ as in [Kottwitz and Shelstad 1999, Section 2.1]; see [Langlands and Shelstad 1987, Section 1.2]. We refer to H as an elliptic endoscopic group for G. We will always be in the case when $\mathcal{H}$ can be taken to be the L-group of H. Our notation for such a datum is usually $\mathfrak{e} = (G^{\mathfrak{e}}, {}^LG^{\mathfrak{e}}, s^{\mathfrak{e}}, \xi^{\mathfrak{e}})$. The set $\mathcal{E}_{\mathrm{ell}}(G)$ always contains a unique element $\mathfrak{e}_0$ whose endoscopic group is a quasisplit inner form of G. Write $\mathcal{E}^{<}_{\mathrm{ell}}(G)$ for the complement $\mathcal{E}_{\mathrm{ell}}(G)\backslash\{\mathfrak{e}_0\}$. Every endoscopic group in $\mathcal{E}^{<}_{\mathrm{ell}}(G)$ has strictly lower semisimple rank than G.

The cyclotomic character has Hodge–Tate weight -1 in our convention.

2. Weak transfer

2.1. ***Classical groups.*** Let $m, n \in \mathbb{Z}_{>0}$. We introduce the quasisplit classical groups Sp_{2n}, SO_{2n+1}, SO^{η}_{2n}, and U_n, naturally sitting inside (the restriction of scalars of) general linear group GL_m. (Compare with [Arthur 2013, Chapters 1 and 9] and [Waldspurger 2010, Section 1].) For unitary groups, we write N instead of m in anticipation of Section 2.2.

Define antidiagonal matrices J_m, $J_m^* \in \mathrm{GL}_m(\mathbb{Z})$ and $J'_{2n} \in \mathrm{GL}_{2n}(\mathbb{Z})$ as follows:

$$J_m = \begin{pmatrix} & & & 1 \\ & & 1 & \\ & \iddots & & \\ 1 & & & \end{pmatrix}, \quad J_m^* = \begin{pmatrix} & & & 1 \\ & & -1 & \\ & \iddots & & \\ (-1)^{m-1} & & & \end{pmatrix}, \quad J'_{2n} = \begin{pmatrix} & -J_n \\ J_n & \end{pmatrix}.$$

When $m = 2n$, let $\eta : \Gamma_{F_\eta/F} \to \{\pm 1\}$ be a faithful character. (So F_η/F is a quadratic extension if $\eta \neq 1$, and $F_\eta = F$ if $\eta = 1$.) If $\eta = 1$ then set $J_m^\eta := J_m$. If $\eta \neq 1$, choose $\alpha \in \mathcal{O}_F^\times$ whose square roots generate F_η over F. Then define J_{2n}^η from J_{2n} by replacing the 2×2-matrix $\left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$ in the middle with $\left(\begin{smallmatrix} 1 & 0 \\ 0 & -\alpha \end{smallmatrix}\right)$.

Case S. We define the $\mathcal{O}_F$-group schemes

$$G \in \{\mathrm{Sp}_m, \mathrm{O}_m^\eta, \mathrm{O}_m\},$$

with $m = 2n$ in the first two cases, and $m = 2n + 1$ in the last case, by the following formula

$$G := \{g \in \mathrm{GL}_m : {}^t g J g = J\}, \quad J \in \{J'_m, J_m^\eta\}, \text{ respectively,}$$

on $\mathcal{O}_F$-algebra valued points. The connected component of the identity in O_m^η (resp. O_{2n+1}) is denoted by SO_{2n}^η (resp. SO_{2n+1}). By abuse of notation, we still write Sp_{2n}, SO_{2n}^η, and SO_{2n+1} for the F-group schemes obtained by base change. We often omit η in case $\eta = 1$. Each group contains a Borel subgroup B over F: if G is SO_m or Sp_{2n} then B consists of upper triangular matrices in G; if $G = \mathrm{SO}_{2n}^\eta$ with $\eta \neq 1$ then B consists of matrices (g_{ij}) such that $g_{ij} = 0$ if $i > j$ and $(i, j) \neq (n+1, n)$. In the following examples, we make an explicit choice of a maximal torus T in B and describe the character group of T as well as the half sum of positive roots ρ. When A_i are square matrices for $1 \leq i \leq r$, let $\mathrm{diag}(A_1, \ldots, A_r)$ denote the block diagonal matrix.

$G = \mathrm{Sp}_{2n}$. We take $T = \{\mathrm{diag}(t_1, \ldots, t_n, t_n^{-1}, \ldots, t_1^{-1}) : t_1, \ldots, t_n \in \mathbb{G}_m\}$ and use the coordinates to identify $X^*(T) = \mathbb{Z}^n$ with trivial Γ_F-action. We have the Weyl group $\Omega = \{\pm 1\}^n \rtimes S_n$, where $(\epsilon_1, \ldots, \epsilon_n) \in \{\pm 1\}^n$ acts on $(a_i) \in X^*(T)$ by sending each a_i to $a_i^{\epsilon_i}$, and S_n acts by permuting $a_1, \ldots, a_n$. By computation $\rho = (n, n-1, \ldots, 2, 1)$.

$G = \mathrm{SO}_{2n}^\eta$ (*allowing* $\eta = 1$). Take $T = \{\mathrm{diag}(t_1, \ldots, t_{n-1}, s, t_{n-1}^{-1}, \ldots, t_1^{-1}) : t_1, \ldots, t_{n-1} \in \mathbb{G}_m, s \in \mathrm{SO}_2^\eta\}$. Using b as the last coordinate we identify $X^*(T) = \mathbb{Z}^n$, with Γ_F acting through η on the last coordinate as $\{\pm 1\}$. The Weyl group Ω is the index two subgroup of $\{\pm 1\}^n \rtimes S_n$ consisting of $(\epsilon_1, \ldots, \epsilon_n, \sigma)$ such that $\prod_{i=1}^n \epsilon_i = 1$. Each element of Ω acts on $\mathbb{Z}^n$ in the same way as in the Sp_{2n}-case. We have $\rho = (n-1, n-2, \ldots, 1, 0)$.

$G = \mathrm{SO}_{2n+1}$. Here $T = \{\mathrm{diag}(t_1, \dots, t_n, 1, t_n^{-1}, \dots, t_1^{-1}) : t_1, \dots, t_n \in \mathbb{G}_m\}$ and $X^*(T) = \mathbb{Z}^n$ with trivial Γ_F-action. The Weyl group $\Omega = \{\pm 1\}^n \rtimes S_n$ acts on $X^*(T)$ in the same way as above, and $\rho = \frac{1}{2}(2n-1, 2n-3, \dots, 3, 1)$.

For each G the choice of (B, T) as above extends to an F-pinning (a.k.a. F-splitting, see [Kottwitz and Shelstad 1999, Section 1.2]). The Langlands dual groups $\widehat{G}$, as reductive groups over $\mathbb{C}$, are described as $\widehat{\mathrm{Sp}_{2n}} = \mathrm{SO}_{2n+1}$, $\widehat{\mathrm{SO}^\eta_{2n}} = \mathrm{SO}_{2n}$, and $\widehat{\mathrm{SO}_{2n+1}} = \mathrm{Sp}_{2n}$, equipped with pinnings for $\widehat{G}$ chosen in the same way as for G. The L-action of Γ_F on $\widehat{G}$ is trivial when G is the split group Sp_{2n}, SO_{2n}, or SO_{2n+1}, whereas the action for $G = \mathrm{SO}^\eta_{2n}$ with $\eta \neq 1$ factors through $\mathrm{Gal}(F_\eta/F)$ with the nontrivial element acts as the outer automorphism $\hat{\theta}^\circ : g \mapsto \vartheta g \vartheta^{-1}$ on SO_{2n}, where

$$\vartheta = \mathrm{diag}\left(I_{n-1}, \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, I_{n-1}\right) \in \mathrm{SO}_{2n}(\mathbb{C}).$$

Set $F' := F$ unless $G = \mathrm{SO}^\eta_{2n}$, in which case $F' := F_\eta$, so that the Γ_F-action factors through $\Gamma_{F'/F}$. Then the F'/F-form of the L-group ${}^L G_{F'/F} = \widehat{G} \rtimes \Gamma_{F'/F}$ is given as follows; we will often omit the subscript F'/F:

$${}^L\,\mathrm{Sp}_{2n} = \mathrm{SO}_{2n+1}, \quad {}^L\,\mathrm{SO}^\eta_{2n} = \begin{cases} \mathrm{O}_{2n}, & \eta \neq 1, \\ \mathrm{SO}_{2n}, & \eta = 1, \end{cases} \quad {}^L\,\mathrm{SO}_{2n+1} = \mathrm{Sp}_{2n},$$

where ${}^L\,\mathrm{SO}^\eta_{2n} = \mathrm{O}_{2n}$ when $\eta \neq 1$ by sending the nontrivial element of $\mathrm{Gal}(F_\eta/F)$ to ϑ.

Endoscopic groups $G^{\mathfrak{e}}$ in $\mathcal{E}_{\mathrm{ell}}(G)$ have the following forms, where $0 \le n' \le n$ and $\eta', \eta'_1, \eta'_2 : \Gamma_F \to \{\pm 1\}$ are continuous characters, understanding that $\eta \neq 1$ (resp. $\eta = 1$) in any factor of the form SO^η_2 (resp. SO^η_0) in the list:

- $G = \mathrm{Sp}_{2n}$: $G^{\mathfrak{e}} = \mathrm{SO}^{\eta'}_{2n'} \times \mathrm{Sp}_{2n-2n'}$.
- $G = \mathrm{SO}^\eta_{2n}$: $G^{\mathfrak{e}} = \mathrm{SO}^{\eta'_1}_{2n'} \times \mathrm{SO}^{\eta'_2}_{2n-2n'}$, $\eta'_1\eta'_2 = \eta$.
- $G = \mathrm{SO}_{2n+1}$: $G^{\mathfrak{e}} = \mathrm{SO}_{2n'+1} \times \mathrm{SO}_{2n+1-2n'}$.

There is redundancy in the second and third items, which can be removed by imposing $n' \le \lfloor n/2 \rfloor$; see [Arthur 2013, Section 1.2] or [Waldspurger 2010, Section 1.8] for a description of full endoscopic data.

Case U. In this case, let E be a quadratic extension of F. Write c for the nontrivial element in $\mathrm{Gal}(E/F)$. Define U_N as an $\mathcal{O}_F$-group scheme by

$$\mathrm{U}_N := \{g \in \mathrm{Res}_{\mathcal{O}_E/\mathcal{O}_F} \mathrm{GL}_N : {}^t g J^*_N c(g) = J^*_N\}$$

on $\mathcal{O}_F$-algebra valued points. Again we still write U_N for $\mathrm{U}_N \times_{\mathcal{O}_F} F$. This group contains a Borel subgroup B (resp. a maximal torus T) over F consisting of upper triangular (resp. diagonal) matrices in U_N so that

$$T = \{(t_1, \dots, t_N) : t_i \in \mathrm{Res}_{E/F}\,\mathbb{G}_m, t_i \cdot c(t_{N+1-i}) = 1, i = 1, \dots, N\}.$$

By fixing an F-algebra embedding $\tau_0 : E \hookrightarrow \overline{F}$, we obtain a projection $(\mathrm{Res}_{E/F}\mathbb{G}_m)_{\overline{F}} \to \mathbb{G}_{m,\overline{F}}$ induced by $E \otimes_F \overline{F} \to \overline{F}$, $a \otimes b \mapsto \tau(a)b$, thereby $T_{\overline{F}} \cong \mathbb{G}^N_{m,\overline{F}}$. This leads to an identification

$$X^*(T) = X_*(\hat{T}) = \mathbb{Z}^N \quad \text{via} \quad \tau_0,$$

with the Γ_F-action factoring through $\Gamma_{E/F}$, and $c \in \Gamma_{E/F}$ acts as $(a_i) \mapsto (-a_{N+1-i})$. (If $\tau_0 c$ was used instead of τ_0, then the identification changes by $(a_i) \mapsto (-a_{N+1-i})$.) We compute $\rho = \left(\frac{1}{2}(N-1), \frac{1}{2}(N-3), \dots, \frac{1}{2}(1-N)\right)$. The above choice of (B, T) extends to an F-pinning.

The map τ_0 induces a projection $(\mathrm{Res}_{E/F}\,\mathrm{GL}_N)_{\overline{F}} \to \mathrm{GL}_{N,\overline{F}}$ inducing $\mathrm{U}_{N,\overline{F}} \cong \mathrm{GL}_{N,\overline{F}}$ and also $\widehat{\mathrm{U}}_N \cong \mathrm{GL}_N$ as a complex reductive group. The standard pinning for GL_N is carried over to a pinning for $\widehat{\mathrm{U}}_N$. The L-action of Γ_F, factoring through $\Gamma_{E/F}$, is given by $c \in \Gamma_{E/F}$ acting as $\hat{\theta}(g) := J_N^{*\,t}g^{-1}(J_N^*)^{-1}$ for $g \in \hat{\mathrm{U}}_N \cong \mathrm{GL}_N$. This determines the structure of the L-group:

$$\tilde{\xi}_0 : {}^L\mathrm{U}_N \cong \mathrm{GL}_N \rtimes \Gamma_F, \quad {}^L(\mathrm{U}_N)_{E/F} \cong \mathrm{GL}_N \rtimes \Gamma_{E/F} \quad \text{via } \tau_0.$$

We also let $\tilde{\xi}_0$ denote either map or the common restriction to the dual group: $\widehat{\mathrm{U}}_N \cong \mathrm{GL}_N$. If τ_0 is replaced with a conjugate embedding $\tau_0 c$, then the above isomorphism is composed with $g \rtimes \gamma \mapsto \hat{\theta}(g) \rtimes \gamma$. Let v be a finite place of F. Recall that $\iota_v : \overline{F} \hookrightarrow \overline{F}_v$ is fixed (Section 1.4), which gives rise to

$$\tau_{0,v} : E \xhookrightarrow{\tau_0} \overline{F} \xhookrightarrow{\iota_v} \overline{F}_v.$$

Write u for the place of E indued by $\overline{F}_v$ via $\tau_{0,v}$. As we did for $\tilde{\xi}_0$, we obtain an isomorphism

$$\tilde{\xi}_u : {}^L(\mathrm{U}_N)_{F_v} = \mathrm{GL}_N \rtimes \Gamma_{F_v} \quad \text{via } \tau_{0,v}.$$

The maps $\tilde{\xi}_0$ and $\tilde{\xi}_u$ fit in a commutative square with the natural embeddings ${}^L(\mathrm{U}_N)_{F_v} \hookrightarrow {}^L\mathrm{U}_N$ and $\mathrm{GL}_N \rtimes \Gamma_{F_v} \hookrightarrow \mathrm{GL}_N \rtimes \Gamma_F$. Similarly, let $\sigma : F \hookrightarrow \mathbb{C}$ be an embedding. Write v for the infinite place of F induced by σ. We have chosen $\iota_\sigma : \overline{F} \hookrightarrow \mathbb{C}$ to extend σ in Section 1.4. Write $\tau_{0,\sigma} := \iota_\sigma \tau_0$. Then we obtain

$$\tilde{\xi}_{\tau_{0,\sigma}} : {}^L(\mathrm{U}_N)_{F_v} = \mathrm{GL}_N \rtimes \Gamma_{F_v} \quad \text{via } \tau_{0,\sigma}.$$

For the embedding $\tau_{0,\sigma}c$ conjugate to $\tau_{0,\sigma}$, we define $\tilde{\xi}_{\tau_{0,\sigma}c}$ to be $\tilde{\xi}_{\tau_{0,\sigma}}$ followed by $g \rtimes \gamma \mapsto \hat{\theta}(g) \rtimes \gamma$. Similarly, if a finite place v splits in E as u and u' then $\tilde{\xi}_{u'}$ is set to be $\tilde{\xi}_u$ composed with $g \rtimes \gamma \mapsto \hat{\theta}(g) \rtimes \gamma$. To sum up, we defined

$$\tilde{\xi}_\tau \text{ for all embeddings } E \hookrightarrow \mathbb{C} \quad \text{and} \quad \tilde{\xi}_u \text{ for all finite places } u \text{ of } E.$$

When v is an infinite place, we also fix an isomorphism $\overline{F}_v \cong \mathbb{C}$ and still write $\tau_{0,v}$ for the composite map $E \hookrightarrow \overline{F}_v \cong \mathbb{C}$. This map induces $\hat{T}_{\tau_{0,v}} \cong \mathbb{G}_m^N$ over $\mathbb{C}$, thus $X_*(\hat{T}_{\tau_{0,v}}) = \mathbb{Z}^N$.

Endoscopic groups in $\mathcal{E}_{\mathrm{ell}}(\mathrm{U}_N)$ have the form $\mathrm{U}_{N_1} \times \mathrm{U}_{N_2}$ for integers $N_1 \geq N_2 \geq 0$ and $N_1 + N_2 = N$. See [Rogawski 1990, Section 4.6]; compare [Waldspurger 2010, Section 1.8] or [Mok 2015, Section 2.4]) for more details on full endoscopic data. We note that the Weil form (rather than the Galois form) of the L-group is needed to describe the L-morphisms in the endoscopic data.

2.2. *Twisted general linear groups.* Consider Cases S and U together. Keep the same E and c as above in Case U; set $E = F$ and $c = 1 \in \mathrm{Gal}(E/F)$ in Case S for uniformity. For $N \in \mathbb{Z}_{\geq 1}$ we introduce the groups

$$\tilde{G}^0(N) := \mathrm{Res}_{E/F}\,\mathrm{GL}_N \quad \text{and} \quad \tilde{G}(N) := \tilde{G}^0(N) \rtimes \langle\theta\rangle,$$

where $\langle\theta\rangle$ is an order 2 group with θ acting on $\tilde{G}^0(N)$ as $\theta(g) : g \mapsto J_N^{*}\,{}^t c(g)^{-1} (J_N^*)^{-1}$. Fix a standard pinning $(B_N, T_N, \{X_N\})$ of $\tilde{G}^0(N)$, which is stabilized by θ. In particular, T_N is the diagonal maximal torus of $\tilde{G}^0(N)$. Write $G(N) := \tilde{G}^0(N) \rtimes \theta$ for the θ-coset in $\tilde{G}(N)$. We also let $G(N)$ stand for the datum $(\tilde{G}(N), \theta)$ as in [Arthur 2013, page 125]. For simplicity of notation we will often write ${}^L G(N)$ and $\widehat{G(N)}$ for ${}^L\tilde{G}^0(N)$ and $\widehat{\tilde{G}^0(N)}$.

Denote by $\tilde{\mathcal{E}}_{\mathrm{ell}}(N)$ a set of representatives for isomorphism classes of twisted endoscopic data for $(\tilde{G}(N), \theta)$. Each element of $\tilde{\mathcal{E}}_{\mathrm{ell}}(N)$ is represented by a quadruple $\tilde{\mathfrak{e}} = (G^{\tilde{\mathfrak{e}}}, {}^L G^{\tilde{\mathfrak{e}}}, s^{\tilde{\mathfrak{e}}}, \xi^{\tilde{\mathfrak{e}}})$; see [Kottwitz and Shelstad 1999]. By $\tilde{\mathcal{E}}_{\mathrm{sim}}(N)$ we mean the subset of simple twisted endoscopic data in $\tilde{\mathcal{E}}_{\mathrm{ell}}(N)$, i.e., the data where $G^{\tilde{\mathfrak{e}}}$ attains maximal semisimple rank.

We give an explicit parametrization of $\tilde{\mathcal{E}}_{\mathrm{ell}}(N)$ by means of the twisted endoscopic group $G^{\tilde{\mathfrak{e}}}$ following [Arthur 2013, Section 1.2] and [Rogawski 1990, Section 4.7]. For simple endoscopic data we will write G and $\tilde{\xi}$ for $G^{\tilde{\mathfrak{e}}}$ and $\xi^{\tilde{\mathfrak{e}}}$, and describe $\tilde{\xi}$ explicitly.

Case S. The twisted endoscopic groups are parametrized by triples

$$(N_O, N_S, \eta), \quad N_O, N_S \in \mathbb{Z}_{\geq 0},\ N_O + N_S = N,\ N_S \text{ is even},\ \eta : \Gamma_F \to \{\pm 1\},$$

where the continuous character η is trivial if $N_O = 0$, nontrivial if $N_O = 2$, and arbitrary if $N_O > 2$. The corresponding $G^{\tilde{\mathfrak{e}}}$ is $\mathrm{SO}^{\eta}_{N_O} \times \mathrm{SO}_{N_S+1}$ if N is even, and $\mathrm{Sp}_{N_O - 1} \times \mathrm{SO}_{N_S+1}$ if N is odd. In each case, $\xi^{\tilde{\mathfrak{e}}}$ can be described as in [Arthur 2013, page 11]. (If N is odd then η only affects $\xi^{\tilde{\mathfrak{e}}}$, not $G^{\tilde{\mathfrak{e}}}$.)

The triple corresponds to an element of $\tilde{\mathcal{E}}_{\mathrm{sim}}(N)$ precisely when $N_O = 0$ or $N_S = 0$. If $N = 2n$, then we have $(0, N, 1)$ and $(N, 0, \eta)$. In the first case, $G = \mathrm{SO}_{2n+1}$ and

$$\tilde{\xi} : {}^L G = \mathrm{Sp}_{2n} \hookrightarrow \mathrm{GL}_{2n}$$

is the standard embedding, inducing the map on cocharacter groups

$$X_*(\widehat{T}) = \mathbb{Z}^n \to X_*(\widehat{T}_{2n}) = \mathbb{Z}^{2n}, \quad (a_i)_{i=1}^n \mapsto (a_1, \ldots, a_n, -a_n, \ldots, -a_1).$$

The triple $(N, 0, \eta)$ corresponds to $G = \mathrm{SO}_{2n}^{\eta}$ and

$$\tilde{\xi} : {}^L G = \mathrm{O}_{2n} \hookrightarrow \mathrm{GL}_{2n}$$

is again the standard embedding, inducing the map on cocharacter groups

$$X_*(\widehat{T}) = X^*(T) = \mathbb{Z}^n \to X_*(\widehat{T}_{2n}) = \mathbb{Z}^{2n}, \quad (a_i)_{i=1}^n \mapsto (a_1, \dots, a_n, -a_n, \dots, -a_1).$$

Strictly speaking the codomain of $\tilde{\xi}$ is $\mathrm{GL}_{2n} \times \Gamma_{F_\eta/F}$, but the image of $\tilde{\xi}$ in the Galois factor is dictated by the fact that $\tilde{\xi}$ is an L-morphism, so we often omit it from the formula. The same will apply to $\tilde{\xi}$ below when N is odd.

If $N = 2n+1$, simple data correspond to $(N, 0, \eta)$, thus $G = \mathrm{Sp}_{2n}$ and

$$\tilde{\xi} : {}^L G_{F_\eta/F} = \mathrm{SO}_{2n+1} \times \Gamma_{F_\eta/F} \hookrightarrow \mathrm{GL}_{2n+1}$$

given by the standard embedding on SO_{2n+1} and $\eta : \Gamma_{F_\eta/F} \hookrightarrow \{\pm 1\} \subset \mathrm{GL}_{2n+1}$ on the Galois group. The induced map on cocharacters is

$$X_*(\widehat{T}) = \mathbb{Z}^n \to X_*(\widehat{T}_{2n+1}) = \mathbb{Z}^{2n+1}, \quad (a_i)_{i=1}^n \mapsto (a_1, \dots, a_n, 0, -a_n, \dots, -a_1).$$

Case U. The twisted endoscopic groups in $\tilde{\mathcal{E}}_{\mathrm{ell}}(N)$ are parametrized by quadruples

$$(N_1, N_2, \kappa_1, \kappa_2), \quad N_1, N_2 \in \mathbb{Z}_{\geq 0}, N_1 + N_2 = N, \kappa_1, \kappa_2 \in \{\pm 1\},$$

with (κ_1, κ_2) either $(1, -1)$ or $(-1, 1)$ if N is even, and $(1, 1)$ or $(-1, -1)$ if N is odd, modulo the equivalence $(N_1, N_2, \kappa_1, \kappa_2) \sim (N_2, N_1, \kappa_2, \kappa_1)$. (Compare with [Mok 2015, Section 2.4], but beware of a small inaccuracy that the equivalence between endoscopic data is incorrect there.) For each quadruple we have a twisted endoscopic group $G^{\mathfrak{e}} = \mathrm{U}_{N_1} \times \mathrm{U}_{N_2}$, with respect to the same E/F, which is part of a twisted endoscopic datum. We refer to *loc. cit.* for a formula for the L-morphism $\xi^{\tilde{\mathfrak{e}}}$, which depends on κ_1, κ_2.

The subset $\tilde{\mathcal{E}}_{\mathrm{sim}}(N)$ corresponds to quadruples $(N, 0, \kappa_1, \kappa_2)$. Set $\kappa := \kappa_1 \in \{\pm 1\}$. We need not keep track of κ_2 as it is determined by N and κ_1. In both cases the twisted endoscopic group is $G = \mathrm{U}_N$; let $\tilde{\xi}_+, \tilde{\xi}_- : {}^L \mathrm{U}_N \to {}^L \tilde{G}^0(N)$ denote the L-morphisms corresponding to $\kappa = 1, -1$, respectively. Let $\tau_0 : E \hookrightarrow \overline{F}$ be the embedding fixed in Section 2.1. Then $\widehat{G(N)} = \mathrm{GL}_N \times \mathrm{GL}_N$, where the copies of GL_N are indexed by τ_0 and $\tau_0 c$ in the order, and $\Gamma_{E/F}$ acts by permuting the two factors. The "base change" morphism $\tilde{\xi}_+$ is easy to describe

$$\begin{aligned} \tilde{\xi}_+ : {}^L(\mathrm{U}_N)_{E/F} \stackrel{\tau_0}{=} \mathrm{GL}_N \rtimes \Gamma_{E/F} &\to {}^L \tilde{G}^0(N) = (\mathrm{GL}_N \times \mathrm{GL}_N) \rtimes \Gamma_{E/F}, \\ g \rtimes \gamma &\mapsto (g, \hat{\theta}(g)) \rtimes \gamma = (g, J_N^{*\,t} g^{-1} (J_N^*)^{-1}) \rtimes \gamma. \end{aligned} \tag{2.2.1}$$

This map is independent of the choice of τ_0: if τ is replaced with τc, then the first identification is twisted by $g \rtimes \gamma \mapsto \hat{\theta}(g) \rtimes \gamma$ while the second map becomes $g \rtimes \gamma \mapsto (g, \hat{\theta}(g))$ (if the first component is still labeled by τ) so the changes are canceled out, while the last identification is unchanged.

The map $\tilde{\xi}_+$ induces a map on the cocharacter groups

$$X_*(\widehat{T}) \stackrel{\tau}{=} \mathbb{Z}^N \to X_*(\widehat{T}_N) = \mathbb{Z}^N \oplus \mathbb{Z}^N, \quad (a_i) \mapsto ((a_i), (-a_{N+1-i}))$$

in accordance with (2.2.1). Similarly we can describe the map induced by $\tilde{\xi}_+$:

$$X_*(\widehat{T}_\infty) = \bigoplus_\sigma X_*(\widehat{T}) \to X_*(\widehat{T}_{N,\infty}) = \bigoplus_\tau X_*(\widehat{T}_N),$$

where the first sum is over embeddings $\sigma : F \hookrightarrow \mathbb{C}$ and the second over $\tau : E \hookrightarrow \mathbb{C}$. Namely if $(a_i) \in X_*(\hat{T})$ denotes the σ-component, then the image is supported on the $\tau_{0,\sigma}$ and $\tau_{0,\sigma}c$ components on the right, and the map is $(a_i) \mapsto ((a_i), (-a_{N+1-i}))$.

We refer to [Mok 2015, Section 2.4] for a description of $\tilde{\xi}_-$, which will be needed only in a minor way, and leaves it as an exercise to describe the induced map on cocharacter groups. We just remark that $\tilde{\xi}_-$ is not defined on L-groups relative to a Galois extension; we need the Weil form of the L-groups.

2.3. ***Global parameters.*** Keep the notation from the preceding subsection. We introduce (conjugate) self-dual parameters for general linear groups, which will serve as parameters for automorphic representations of classical groups. We are following [Arthur 2013, Section 1.4] in spirit, but our situation is simpler in that we do not need the seed theorems of Arthur (namely [Arthur 2013, Theorems 1.4.1 and 1.4.2]) as we will prove only weak transfers.

For $m \in \mathbb{Z}_{\geq 1}$, let $\Psi_{\text{sim}}(m)$ denote the set of (isomorphism classes of) unitary cuspidal automorphic representations of $G(m, \mathbb{A}_F) = \mathrm{GL}_m(\mathbb{A}_E)$. Write $\Psi(N)$ for the set of formal global parameters

$$\psi = \boxplus_{i \in I} \mu_i \boxtimes \nu_{n_i}, \quad \mu_i \in \Psi_{\text{sim}}(m_i), m_i, n_i \in \mathbb{Z}_{\geq 1}, \tag{2.3.1}$$

where I is a finite index set, ν_{n_i} is an irreducible n_i-dimensional algebraic representation of $\mathrm{SL}_2(\mathbb{C})$, and $\sum_{i \in I} m_i n_i = N$. Given ψ is considered equal to another parameter $\psi' = \boxplus_{i' \in I'} \mu_{i'} \boxtimes \nu_{n'_i}$ if there exists a bijection $f : I \to I'$ such that $\mu_i = \mu_{f(i)}$ and $n_i = n_{f(i)}$ for all $i \in I$.

Given $\mu \in \Psi_{\text{sim}}(m)$, let $\mu^\star := \mu^\vee \circ c \in \Psi_{\text{sim}}(m)$ denote its conjugate-dual. This definition extends to $\Psi(N)$ by setting $\psi^\star := \boxplus_{i \in I} \mu_i^\star \boxtimes \nu_{n_i}$. Put

$$\tilde{\Psi}(N) := \{\psi \in \Psi(N) : \psi^\star = \psi\}.$$

Let S be a finite set of places of F containing all the places of F ramified in E. Write $\Psi^S(N)$ for the subset of $\psi \in \Psi(N)$ which are unramified outside S; the latter means that μ_i are all unramified outside S in (2.3.1). Put $\tilde{\Psi}^S(N) := \tilde{\Psi}(N) \cap \Psi^S(N)$. We define $\mathcal{C}_\infty(N)$ and $\mathcal{C}^S(N)$ to be the sets of $\mathbb{C}$-algebra characters of $\mathfrak{Z}(\tilde{G}^0(N)_\infty)$ and $\mathcal{H}^S_{\text{ur}}(\tilde{G}^0(N))$, respectively. We have a map

$$\psi \in \Psi^S(N) \mapsto (\zeta_{\psi,\infty}, c^S(\psi)) \in \mathcal{C}_\infty(N) \times \mathcal{C}^S(N)$$

defined as follows. Given ψ as in (2.3.1), we have $(\zeta_{\mu_{i,\infty}}, c^S(\mu_i)) \in \mathcal{C}_\infty(m_i) \times \mathcal{C}^S(m_i)$. The block diagonal embedding $\prod_{i\in I}\prod_{j=1}^{n_i} \mathrm{GL}_{m_i} \to \mathrm{GL}_{m_i n_i}$ induces a map

$$\prod_{i\in I}\prod_{j=1}^{n_i}(\mathcal{C}_\infty(m_i)\times\mathcal{C}^S(m_i)) \to \mathcal{C}_\infty(N)\times\mathcal{C}^S(N).$$

We define $(\zeta_{\psi_\infty}, c^S(\psi))$ to be the image of

$$\left(\zeta_{\mu_{i,\infty}} + \frac{n_i+1-2j}{2},\, q_v^{(n_i+1-2j)/2}c^S(\mu_i)\right)_{i\in I, 1\le j\le n_i},$$

where the sum $\zeta_{\mu_{i,\infty}}+a$ with $a\in\mathbb{Q}$ means that the sum is taken in $X_*(\widehat{T}_{m_i})_{\mathbb{C}}/\Omega_{m_i}$, and $a\in\mathbb{Q}=X_*(\mathbb{G}_m)_{\mathbb{Q}}$ embeds into $X_*(\widehat{T}_{m_i})_{\mathbb{Q}}$ via the inclusion of $\mathbb{G}_m = Z(\widehat{\tilde{G}^0(m_i)})^{\Gamma_F}$ in $\widehat{T}_{m_i}$; the product $q_v^b c^S(\psi)$ with $b\in\mathbb{Q}$ is taken in $\widehat{\tilde{G}^0(m_i)}$, where $q_v^b\in\mathbb{G}_m(\mathbb{C})$ is viewed as a central element of the dual group of $\tilde{G}^0(m_i)$. Our definition of $(\zeta_{\psi,\infty}, c^S(\psi))$ is given explicitly such that it is consistent with the local A-parameters at ∞ and finite places away from S obtained from localizing ψ.

2.4. ***Stabilized trace formulas.*** Let G be an inner form of a quasisplit classical group as in Section 2.1. (In fact the discussion below in the untwisted case works for general reductive groups as in the relevant parts of [Arthur 2013, Chapter 3].)

Let us begin by introducing the notion of Hecke types following [Arthur 2013, page 129]. We freely use the notation and the choices made from Section 1.4. Let S be a finite set of places of F containing $S_{\mathrm{bad}}(G)$. Let κ_S^∞ be an open compact subgroup of $\prod_v G(F_v)$, where v runs over finite places in S. Write K^S for the product of hyperspecial subgroups K_v^0 over finite places $v\notin S$, so $\kappa_S^\infty K^S$ is an open compact subgroup of $G(\mathbb{A}_F^\infty)$. Fix a finite set τ_∞ consisting of irreducible representations of a fixed maximal compact subgroup K_∞ of $G_\infty(\mathbb{R}) = \prod_{v\mid\infty} G(F_v)$. The pair $\kappa = (\tau_\infty, \kappa_S^\infty K^S)$ arising this way is called a Hecke type. Write $\mathcal{H}(G)_\kappa$ for the subspace generated by $f = f^\infty f_\infty \in \mathcal{H}(G)$ such that f^∞ is biinvariant under $\kappa_S^\infty K^S$ and such that f_∞ transforms under left and right translations under K_∞ according to representations in τ_∞.

Let $h\in\mathcal{H}_{\mathrm{ur}}^S(G)$ and $z\in\mathfrak{Z}(G_\infty)$. By evaluating $c^S\in\mathcal{C}^S(G)$ and $\zeta\in\mathcal{C}_\infty(G)$ at h and z respectively (see Section 1.4), we obtain the numbers to be denoted by $\widehat{h}(c^S)\in\mathbb{C}$ and $\zeta(z)\in\mathbb{C}$. Moreover h and z act on $\mathcal{H}_{\mathrm{ur}}^S(G)$ and $\mathcal{H}(G_\infty)$, written as $f^S\mapsto h*f^S$ and $f_\infty\mapsto z*f_\infty$, such that for irreducible admissible representations π^S of $G(\mathbb{A}_F^S)$ and π_∞ of $G_\infty(\mathbb{R})$,

$$\pi^S(h*f^S) = \widehat{h}(c^S(\pi^S))\pi^S(f^S), \quad \pi_\infty(z*f_\infty) = \zeta_{\pi_\infty}(z)\pi_\infty(f_\infty). \tag{2.4.1}$$

In particular we have identities by taking the traces of both sides in (2.4.1). The commuting action of (h, z) on $\mathcal{H}^S_{\mathrm{ur}}(G) \times \mathcal{H}(G_\infty)$, again denoted by $*$, obviously extends to $\mathcal{H}(G(\mathbb{A}_F), K^S)$.

Let $t \in \mathbb{R}_{\geq 0}$. Write $I^G_{\mathrm{disc},t}$ for the discrete part of the trace formula, which is an invariant linear form on $\mathcal{H}(G)$. The restriction of $I^G_{\mathrm{disc},t}$ to $\mathcal{H}(G)_\kappa$ decomposes as a finite sum of eigen-linear forms of $\mathcal{H}^S_{\mathrm{ur}}(G)$. Moreover, we can further decompose as a finite sum of eigen-linear forms for the action of $\mathfrak{Z}(G_\infty)$ on $\mathcal{H}(G_\infty)$. Thus we can write

$$I^G_{\mathrm{disc},t}(f) = \sum_{(\zeta, c^S) \in \mathcal{C}_\infty(G) \times \mathcal{C}^S(G)} I^G_{\mathrm{disc},\zeta,c^S}(f), \quad f \in \mathcal{H}(G(\mathbb{A}_F), K^S), \tag{2.4.2}$$

where $I^G_{\mathrm{disc},\zeta,c^S}$ are (ζ, c^S)-eigen-linear forms:

$$I^G_{\mathrm{disc},\zeta,c^S}((h, z) * f) = \widehat{h}(c^S)\zeta(z) I^G_{\mathrm{disc},\zeta,c^S}(f), \quad h \in \mathcal{H}^S_{\mathrm{ur}}(G), z \in \mathfrak{Z}(G_\infty). \tag{2.4.3}$$

The ζ and c^S appearing in (2.4.2) should be thought of as the infinitesimal characters at ∞ and the away-from-S Satake parameters for the automorphic representations contributing to $I_{\mathrm{disc},t}$. For a fixed Hecke type κ, the sum (2.4.2) runs over a finite set depending only on κ and not on $f \in \mathcal{H}(G)_\kappa$ by Harish-Chandra's finiteness theorem.

Note that t is determined by ζ to be the norm of the imaginary part of ζ; see [Arthur 2013, page 123]. That is, for a fixed ζ and c^S, the linear form $I^G_{\mathrm{disc},\zeta,c^S}$ in (2.4.2) is nontrivial for at most one t. Hence the meaning of $I^G_{\mathrm{disc},\zeta,c^S}$ is unambiguous even if we do not include t in the notation.

Write $R^G_{\mathrm{disc},t}$ for the regular representation of $G(\mathbb{A}_F)$ on $L^2_{\mathrm{disc}}([G])$; see Section 1.4. Just like $I^G_{\mathrm{disc},t}$ the invariant distribution $\operatorname{tr} R^G_{\mathrm{disc},t}$ decomposes as

$$\operatorname{tr} R^G_{\mathrm{disc},t}(f) = \sum_{(\zeta, c^s) \in \mathcal{C}_\infty(G) \times \mathcal{C}^S(G)} \operatorname{tr} R^G_{\mathrm{disc},\zeta,c^S}(f), \quad f \in \mathcal{H}(G(\mathbb{A}_F), K^S).$$

To discuss stable distributions, we will only consider G with the following property: for every finite sequence $\mathfrak{e}_i = (G^{\mathfrak{e}}_i, \mathcal{G}^{\mathfrak{e}}_i, s^{\mathfrak{e}}_i, \xi^{\mathfrak{e}}_i)$ indexed by $i = 1, \dots, r$, where $\mathfrak{e}_i$ is an elliptic endoscopic datum for $G^{\mathfrak{e}}_{i-1}$ over F for $2 \leq i \leq r$, we can take $\mathcal{G}^{\mathfrak{e}}_i = {}^L G^{\mathfrak{e}}_i$ for all $1 \leq i \leq r$. (That is, $\mathfrak{e}_i$ is isomorphic to an endoscopic datum whose second entry is given by the L-group of the first entry.) The purpose of the simplifying hypothesis is to dispense with any discussion of z-extensions. This suffices for our needs as the classical groups in Section 2.1 satisfy the condition.

Now we consider elliptic endoscopic data $\mathfrak{e} = (G^{\mathfrak{e}}, \mathcal{G}^{\mathfrak{e}}, s^{\mathfrak{e}}, \xi^{\mathfrak{e}})$ for G over F. Denote by $f^{\mathfrak{e}} \in \mathcal{H}(G^{\mathfrak{e}}(\mathbb{A}_F))$ a Langlands–Shelstad transfer of f. Arthur inductively defined stable linear forms $S^{\mathfrak{e}}_{\mathrm{disc},t} = S^{G^{\mathfrak{e}}}_{\mathrm{disc},t} : \mathcal{H}(G^{\mathfrak{e}}) \to \mathbb{C}$ for each $\mathfrak{e}$ satisfying the fundamental identity

$$I^G_{\mathrm{disc},t}(f) = \sum_{\mathfrak{e}\in\mathcal{E}_{\mathrm{ell}}(G)} \iota(\mathfrak{e}) S^{\mathfrak{e}}_{\mathrm{disc},t}(f^{\mathfrak{e}}), \tag{2.4.4}$$

where $\iota(\mathfrak{e}) \in \mathbb{Q}_{>0}$ is an explicit constant. For quasisplit $G = G^{\mathfrak{e}_0}$, the equality should be viewed as an inductive definition of $S^G_{\mathrm{disc},t}$; the inductive procedure is based on the fact that the semisimple rank of $G^{\mathfrak{e}}$ is less than that of G for $\mathfrak{e} \in \mathcal{E}^{<}_{\mathrm{ell}}(G)$. The role of the stabilization of the trace formula is to tell us that the inductive definition of $S^G_{\mathrm{disc},t}$ indeed yields a stable linear form. If G is not quasisplit then both sides of (2.4.4) are a priori defined, and the content of the stabilization is that the equality holds in (2.4.4). See the explanation between (3.2.3) and (3.2.4) in [Arthur 2013] for more details.

The transfer $f^{\mathfrak{e}}$ has trivial stable orbital integrals unless $S \supset S_{\mathrm{bad}}(G^{\mathfrak{e}})$, which we assume from now. In particular if $f \in \mathcal{H}(G(\mathbb{A}_F), K^S)$ then $f^{\mathfrak{e}} \in \mathcal{H}(G^{\mathfrak{e}}(\mathbb{A}_F), K^{\mathfrak{e},S})$, where $K^{\mathfrak{e},S}$ is the product of fixed hyperspecial subgroups of $G^{\mathfrak{e}}(F_v)$ over $v \notin S$. Based on (2.4.2) and (2.4.4), we can adapt the argument from [Arthur 2013, Lemma 3.3.1] to decompose $S^{\mathfrak{e}}_{\mathrm{disc},t}$ into stable linear forms

$$S^{\mathfrak{e}}_{\mathrm{disc},t}(f^{\mathfrak{e}}) = \sum_{(\zeta',c'^{,S})\in\mathcal{C}_\infty(G^{\mathfrak{e}})\times\mathcal{C}^S(G^{\mathfrak{e}})} S^{\mathfrak{e}}_{\mathrm{disc},\zeta',c'^{,S}}(f^{\mathfrak{e}}), \quad f \in \mathcal{H}(G^{\mathfrak{e}}(\mathbb{A}_F), K^{\mathfrak{e},S}),$$

such that each $S^{\mathfrak{e}}_{\zeta',c'^{,S}}$ satisfies the analogue of (2.4.3). If G is quasisplit, then this applies in particular to $G^{\mathfrak{e}} = G$, that is, we have a stable linear form $S^G_{\mathrm{disc},\zeta,c^S} : \mathcal{H}(G(\mathbb{A}_F), K^S) \to \mathbb{C}$ for (ζ, S) as before. Given $(\zeta, c^S) \in \mathcal{C}_\infty(G) \times \mathcal{C}^S(G)$, define

$$S^{\mathfrak{e}}_{\mathrm{disc},\zeta,c^S} := \begin{cases} \sum_{(\zeta',c'^{,S})\mapsto(\zeta,c^S)} S^{\mathfrak{e}}_{\mathrm{disc},\zeta',c'^{,S}} & \text{if } S \supset S_{\mathrm{bad}}(G^{\mathfrak{e}}), \\ 0 & \text{otherwise.} \end{cases}$$

where the sum is taken over the pairs such that $\zeta' \mapsto \zeta$ and $c'^{,S} \mapsto c^S$ under the natural maps $\mathcal{C}_\infty(G^{\mathfrak{e}}) \to \mathcal{C}_\infty(G)$ and $\mathcal{C}^S(G^{\mathfrak{e}}) \to \mathcal{C}^S(G)$ induced by $\xi^{\mathfrak{e}}$. Then we have a refinement of (2.4.4) as in [Arthur 2013, Lemma 3.3.1]:

$$I^G_{\mathrm{disc},\zeta,c^S}(f) = \sum_{\mathfrak{e}\in\mathcal{E}_{\mathrm{ell}}(G)} \iota(\mathfrak{e}) S^{\mathfrak{e}}_{\mathrm{disc},\zeta,c^S}(f^{\mathfrak{e}}). \tag{2.4.5}$$

More precisely, the refinement by c^S is done in [loc. cit.] but not by infinitesimal characters. The argument of [loc. cit.] based on multipliers works in the same way to give refinement by ζ as long as the archimedean transfer is compatible with infinitesimal characters; such compatibility is stated and proved in either of [Mezo 2013, Lemma 24] and [Mœglin and Waldspurger 2016a, I.2.8. Corollary], including the twisted case. This point is also explained in [Taïbi 2019, page 867].

The discussion so far can be adapted to the twisted case, as this case is covered in [Arthur 2013, Sections 3.1–3.3]. For the twisted group $\tilde{G}(N)$ introduced in

Section 2.1, denote by $I^{G(N)}_{\mathrm{disc},t}$ the twisted invariant trace formula and by $\tilde{\mathcal{E}}_{\mathrm{ell}}(N)$ a set of representatives for isomorphism classes of twisted endoscopic data. Each $\tilde{\mathfrak{e}} \in \tilde{\mathcal{E}}_{\mathrm{ell}}(N)$ is again represented by a quadruple $(G^{\tilde{\mathfrak{e}}}, {}^LG^{\tilde{\mathfrak{e}}}, s^{\tilde{\mathfrak{e}}}, \xi^{\tilde{\mathfrak{e}}})$, where $G^{\tilde{\mathfrak{e}}}$ is a product of one or two classical groups as listed in Section 2.2.

Recall that we defined $\mathcal{C}_\infty(N)$ and $\mathcal{C}^S(N)$ in Section 2.3. Put $K(N)^S \subset \tilde{G}^0(N)(\mathbb{A}^S_F)$ for the product of hyperspecial subgroups coming from the obvious integral model of $\tilde{G}^0(N)$ over $\mathcal{O}_F$. We have $h \in \mathcal{H}^S_{\mathrm{ur}}(\tilde{G}^0(N))$ and $z \in \mathfrak{Z}(\tilde{G}^0(N)_\infty)$ act on $\mathcal{H}(G(N, \mathbb{A}^S_F), K(N)^S)$ and $\mathcal{H}(G(N)_\infty)$, respectively, such that the analogue of (2.4.1) holds for representations of $\tilde{G}(N, \mathbb{A}^S_F)$ and $\tilde{G}(N)_\infty$. The decomposition (2.4.2) admits a twisted analogue

$$I^{G(N)}_{\mathrm{disc},t}(f) = \sum_{(\tilde{\zeta}, \tilde{c}^S) \in \mathcal{C}_\infty(N) \times \mathcal{C}^S(N)} I^{G(N)}_{\mathrm{disc},\tilde{\zeta},\tilde{c}^S}(f), \quad f \in \mathcal{H}(G(N, \mathbb{A}_F), K(N)^S),$$

where each $I^{G(N)}_{\mathrm{disc},\tilde{\zeta},\tilde{c}^S}$ is an invariant linear form on $\mathcal{H}(G(N))$ satisfying the eigenproperty analogous to (2.4.3). As before, $I^{G(N)}_{\mathrm{disc},\tilde{\zeta},\tilde{c}^S}$ is nontrivial for at most one t, so there is no danger if t is omitted in the subscript.

Provided that $S \supset S_{\mathrm{bad}}(G^{\tilde{\mathfrak{e}}})$, the L-morphism $\xi^{\tilde{\mathfrak{e}}} : {}^LG^{\tilde{\mathfrak{e}}} \to {}^L\tilde{G}^0(N)$ induces maps $\mathcal{C}_\infty(G^{\tilde{\mathfrak{e}}}) \to \mathcal{C}_\infty(N)$ and $\mathcal{C}^S(G^{\tilde{\mathfrak{e}}}) \to \mathcal{C}^S(N)$. Thereby we put, for each $(\tilde{\zeta}, \tilde{c}^S) \in \mathcal{C}_\infty(N) \times \mathcal{C}^S(N)$,

$$S^{\tilde{\mathfrak{e}}}_{\mathrm{disc},\tilde{\zeta},\tilde{c}^S} := \sum_{(\zeta, c^S) \mapsto (\tilde{\zeta}, \tilde{c}^S)} S^{\tilde{\mathfrak{e}}}_{\mathrm{disc},\zeta,c^S},$$

as a stable linear form on $\mathcal{H}(G^{\tilde{\mathfrak{e}}})$. If $S \not\supset S_{\mathrm{bad}}(G^{\tilde{\mathfrak{e}}})$ then set $S^{\tilde{\mathfrak{e}}}_{\mathrm{disc},\tilde{\zeta},\tilde{c}^S} := 0$.

The stabilization of the twisted trace formula due to Moeglin and Waldspurger [2016b, X.8.1] shows that, if $f^{\tilde{\mathfrak{e}}}$ denotes a Langlands–Shelstad–Kottwitz transfer of $f \in \mathcal{H}(G(N))$ then the twisted analogue of (2.4.4) holds:

$$I^{G(N)}_{\mathrm{disc},t}(f) = \sum_{\tilde{\mathfrak{e}} \in \tilde{\mathcal{E}}_{\mathrm{ell}}(N)} \iota(\tilde{\mathfrak{e}}) S^{\tilde{\mathfrak{e}}}_{\mathrm{disc},t}(f^{\tilde{\mathfrak{e}}}), \tag{2.4.6}$$

where $\iota(\tilde{\mathfrak{e}}) \in \mathbb{Q}_{>0}$ is an explicit constant. For $(\tilde{\zeta}, \tilde{c}^S)$ as above, we refine the preceding formula again by [Arthur 2013, Lemma 3.3.1] (see the paragraph below (2.4.5)):

$$I^{G(N)}_{\mathrm{disc},\tilde{\zeta},\tilde{c}^S}(f) = \sum_{\tilde{\mathfrak{e}} \in \tilde{\mathcal{E}}_{\mathrm{ell}}(N)} \iota(\tilde{\mathfrak{e}}) S^{\tilde{\mathfrak{e}}}_{\mathrm{disc},\tilde{\zeta},\tilde{c}^S}(f^{\tilde{\mathfrak{e}}}). \tag{2.4.7}$$

2.5. ***Weak transfer for classical groups.*** Let G^* be a quasisplit classical group as in Case S or U of Section 2.1. Let $\tilde{\xi} : {}^LG^* \to {}^L\tilde{G}^0(N)$ be the L-morphism such that G^* and $\tilde{\xi}$ constitute a simple twisted endoscopic group for $(\tilde{G}(N), \theta)$ as in Section 2.2. Let $(G, \mathfrak{i})$ be an inner twist of G^* over F (Section 1.4).

Theorem 2.5.1 (quasisplit case). *Assume (H1) in Section 1.1 and let $G = G^*$. Fix a finite set $S \supset S_{\mathrm{bad}}(G)$:*

(1) *For $(\zeta, c^S) \in \mathcal{C}_\infty(G) \times \mathcal{C}^S(G)$ write $(\tilde{\zeta}, \tilde{c}^S) \in \mathcal{C}_\infty(N) \times \mathcal{C}^S(N)$ for the image of (ζ, c^S) under $\tilde{\xi}$. Unless $(\tilde{\zeta}, \tilde{c}^S) = (\zeta_{\psi,\infty}, c^S(\psi))$ for some $\psi \in \tilde{\Psi}^S(N)$,*

$$\operatorname{tr} R^G_{\mathrm{disc},\zeta,c^S}(f) = I^G_{\mathrm{disc},\zeta,c^S}(f) = S^G_{\mathrm{disc},\zeta,c^S}(f) = 0, \quad f \in \mathcal{H}(G(\mathbb{A}_F), K^S).$$

(2) *We have a $G(\mathbb{A}_F)$-equivariant decomposition*

$$L^2_{\mathrm{disc}}([G])^{S-\mathrm{ur}} = \bigoplus_{\psi} \bigoplus_{(\zeta,c^S)\mapsto(\zeta_{\psi,\infty},c^S(\psi))} L^2_{\mathrm{disc},\zeta,c^S}([G])$$

where the first sum runs over $\psi \in \tilde{\Psi}^S(N)$, and the second over $(\zeta, c^S) \in \mathcal{C}_\infty(G) \times \mathcal{C}^S(G)$ which map to $(\zeta_{\psi,\infty}, c^S(\psi))$ under $\tilde{\xi}$. (*See Section 2.3 for the notation.*)

This theorem corresponds to [Arthur 2013, Proposition 3.4.1, Corollary 3.4.3]. Arthur's main global theorems (Section 1.5 therein) show that only a proper subset of $\tilde{\Psi}^S(N)$ contributes in (i) and (ii), consisting of the ones coming from square-integrable parameters of G. The soft argument here does not narrow down the set of ψ as much. Theorem 2.5.1 is proven essentially in the same way as [Arthur 2013, Proposition 3.4.1, Corollary 3.4.3]. We give some details for the convenience of the reader, taking for granted the key input [Arthur 2013, Proposition 3.5.1] on vanishing.

Proof. Assume that $(\tilde{\zeta}, \tilde{c}^S) \neq (\zeta_{\psi,\infty}, c^S(\psi))$ for any $\psi \in \tilde{\Psi}^S(N)$. Let us show (i) and (ii) by induction on N.

Let us check (i) and (ii) when G is a torus; this serves as the base case. Concretely $G = \mathrm{SO}_2^\eta$ (allowing $\eta = 1$) in Case S, and $G = \mathrm{U}_1$ in Case U. Since the two cases are similar, we only consider the latter case. Then $L^2_{\mathrm{disc}}([G])^{S-\mathrm{ur}} = \bigoplus_\chi \chi$, where $\chi : \mathrm{U}_1(\mathbb{A}_F)\backslash \mathrm{U}_1(\mathbb{A}_F) \to \mathbb{C}^\times$ is an automorphic character unramified outside S. This matches the decomposition on the right-hand side of (ii) since each χ determines a unique conjugate self-dual Hecke character $\psi : E^\times\backslash\mathbb{A}_E^\times \to \mathbb{C}^\times$ by $\psi(x) = \chi(x/x^c)$ and a unique pair (ζ, c^S) recording the infinitesimal character and the Satake parameter of χ. Turning to the displayed formula of (i), we see that the first equality holds because a torus has no proper parabolic subgroup, and that the second equality holds because a torus permits no elliptic endoscopic data other than the tautological one. Now the vanishing of the quantities in (i) follows from the decomposition of (ii).

Now we proceed with the induction hypothesis- suppose that (i) and (ii) are known for all quasisplit classical groups which are simple twisted endoscopic groups of $G(N')$ for all $N' < N$ and that G is a simple twisted endoscopic group for $G(N)$. (Here $N > 1$.)

Recall that $I^G_{\mathrm{disc},t} - \mathrm{tr}\, R^G_{\mathrm{disc},t}$ is by definition a linear combination of traces of induced representations from discrete automorphic representations π_M on proper Levi subgroups M of G. So the same is true for $I^G_{\mathrm{disc},\zeta,c^S} - \mathrm{tr}\, R^G_{\mathrm{disc},\zeta,c^S}$. Hence, if the latter were nonzero, then there exists a proper Levi M of G such that (ζ, c) is the image of $\mathfrak{c} = (\zeta_M, c^S_M) \in \mathcal{C}_\infty(M) \times \mathcal{C}^S(M)$ associated with some discrete automorphic representation π_M of $M(\mathbb{A}_F)$. We can write $M = M_h \times M_l$ with M_h a classical group, where M_h is realized as a twisted endoscopic group for $G(N - 2N')$, and $M_l = G(N')$ with $N' < N$. According to $M = M_h \times M_l$, we decompose $\mathfrak{c} = (\mathfrak{c}_h, \mathfrak{c}_l)$. By induction hypothesis for M_h, we have $\mathfrak{c}_h$ map to $(\zeta_{\psi_h,\infty}, c^S(\psi_h))$ for some $\psi_h \in \tilde{\Psi}(N - 2N')$. On the other hand, since the L^2-discrete spectrum of M_l is completely accounted for by $\Psi(N')$ thanks to [Mœglin and Waldspurger 1989] (see [Arthur 2013, pages 23–25] for explanation), we have $\mathfrak{c}_l = (\zeta_{\psi_l,\infty}, c^S(\psi_l))$ for some $\psi_l \in \Psi(N')$. Since (ζ, c) is the image of $(\mathfrak{c}_h, \mathfrak{c}_l)$ under parabolic induction, we see that $(\tilde{\zeta}, \tilde{c}^S) = (\zeta_{\psi,\infty}, c^S(\psi))$ for $\psi = \psi_h \boxplus \psi_l \boxplus \psi_l^\star \in \tilde{\Psi}(N)$. This is a contradiction. We conclude that

$$I^G_{\mathrm{disc},\zeta,c^S}(f) = \mathrm{tr}\, R^G_{\mathrm{disc},\zeta,c^S}(f), \quad f \in \mathcal{H}(G(\mathbb{A}_F), K^S). \tag{2.5.1}$$

Now $I^G_{\mathrm{disc},\zeta,c^S} - S^G_{\mathrm{disc},\zeta,c^S}$ is a linear combination of $S^{\mathfrak{e}}_{\mathrm{disc},\zeta,c^S}$ over $\mathfrak{e} \in \mathcal{E}^<_{\mathrm{ell}}(G)$. If the difference were nonzero, then for some $\mathfrak{e}$,

$$S^{\mathfrak{e}}_{\mathrm{disc},\zeta,c^S} = \sum_{(\zeta',c'^{,S}) \mapsto (\zeta,c^S)} S^{\mathfrak{e}}_{\mathrm{disc},\zeta',c'^{,S}}$$

is nontrivial. Since $G^{\mathfrak{e}}$ is a product of quasisplit classical groups G_1 and G_2 of lower rank (see Section 2.1), by arguing as in the preceding paragraph based on the induction hypothesis for G_1 and G_2, we reach a similar contradiction. (The difference is that there is no general linear factor in G and that the role of parabolic induction is played by the endoscopic transfer via $\xi^{\mathfrak{e}}$.) Hence

$$I^G_{\mathrm{disc},\zeta,c^S}(f) = S^G_{\mathrm{disc},\zeta,c^S}(f), \quad f \in \mathcal{H}(G(\mathbb{A}_F), K^S). \tag{2.5.2}$$

By the initial hypothesis, $I^{G(N)}_{\mathrm{disc},\tilde{\zeta},\tilde{c}^S} = 0$. Applying (2.4.7), (2.5.1) and (2.5.2), we obtain

$$0 = I^{G(N)}_{\mathrm{disc},\tilde{\zeta},\tilde{c}^S}(\tilde{f}) = \sum_{\tilde{\mathfrak{e}} \in \tilde{\mathcal{E}}_{\mathrm{ell}}(N)} \iota(\tilde{\mathfrak{e}})\, \mathrm{tr}\, R^{G^{\tilde{\mathfrak{e}}}}_{\mathrm{disc},\tilde{\zeta},\tilde{c}^S}(\tilde{f}^{\tilde{\mathfrak{e}}}). \tag{2.5.3}$$

The sum runs over the set of $\tilde{\mathfrak{e}}$ such that $G^{\tilde{\mathfrak{e}}}$ is unramified outside S; thus it is a finite sum. Each $\mathrm{tr}\, R^{G^{\tilde{\mathfrak{e}}}}_{\mathrm{disc},\tilde{\zeta},\tilde{c}^S}$ is a positive linear combination of traces of finitely many discrete automorphic representations $\pi^{\tilde{\mathfrak{e}}}$ of $G^{\tilde{\mathfrak{e}}}(\mathbb{A}_F)$. If $\tilde{f}$ is chosen from the Hecke algebra on $G(N)$ of a fixed Hecke type κ then each $\tilde{f}^{\tilde{\mathfrak{e}}}$ belongs to the Hecke algebra on $G^{\tilde{\mathfrak{e}}}$ of a Hecke type $\kappa^{\tilde{\mathfrak{e}}}$ determined by κ. Thus the set of contributing $\pi^{\tilde{\mathfrak{e}}}$ is contained in a finite set depending only on κ, by the condition that $\pi^{\tilde{\mathfrak{e}}}$ should be unramified outside S and that the components of $\pi^{\tilde{\mathfrak{e}}}$ at S should have finitely many types dictated by $\kappa^{\tilde{\mathfrak{e}}}$. (The discussion of this paragraph is based on the explanation between (3.4.11) and (3.4.13) of [Arthur 2013]. The two key facts are that a compatible family therein arises exactly from an element of the Hecke algebra on $G(N)$ and that a compatible family always has a Hecke type.)

The preceding paragraph tells us that Arthur's vanishing result [2013, Proposition 3.5.1] applies to (2.5.3). As a result, every summand in (2.5.3) is identically zero. In particular this is true for $G^{\tilde{\mathfrak{e}}} = G$, namely $\operatorname{tr} R^G_{\mathrm{disc},\tilde{\zeta},\tilde{c}^S}$ is an empty linear combination. That is, $\operatorname{tr} R^G_{\mathrm{disc},\tilde{\zeta},\tilde{c}^S}(f) = 0$ for all f. This completes the proof of (i) in light of (2.5.1) and (2.5.2).

Part (ii) follows immediately from (i) since $\operatorname{tr} R^G_{\mathrm{disc},\zeta,c^S} = 0$, which implies $L^2_{\mathrm{disc},\zeta,c^S}([G]) = 0$, unless (ζ, c^S) maps to $(\zeta_{\psi,\infty}, c^S(\psi))$ for some $\psi \in \tilde{\Psi}^S(N)$. □

Theorem 2.5.2 (general case). *Assume (H1). Let* $(G, \mathfrak{i})$ *be an inner twist of* G^* *over* F. *For each* $\zeta \in \mathcal{C}_\infty(G)$ *and* $c^S \in \mathcal{C}^S(G)$,

$$\operatorname{tr} R^G_{\mathrm{disc},\zeta,c^S}(f) = I^G_{\mathrm{disc},\zeta,c^S}(f) = 0, \qquad f \in \mathcal{H}(G(\mathbb{A}_F), K^S),$$

unless ξ *sends* (ζ, c^S) *to* $(\zeta_{\psi,\infty}, c^S(\psi))$ *for some* $\psi \in \tilde{\Psi}(N)$. *There is a* $G(\mathbb{A}_F)$-*equivariant decomposition*

$$L^2_{\mathrm{disc}}([G])^{S-\mathrm{ur}} = \bigoplus_{\psi} \bigoplus_{(\zeta,c^S)\mapsto(\zeta_{\psi,\infty},c^S(\psi))} L^2_{\mathrm{disc},\zeta,c^S}([G]),$$

where the sums run over $\psi \in \tilde{\Psi}^S(N)$ *and* $(\zeta, c^S) \in \mathcal{C}_\infty(G) \times \mathcal{C}^S(G)$ *such that* $\xi((\zeta, c^S)) = (\zeta_{\psi,\infty}, c^S(\psi))$.

Proof. We induct on N as in the proof of Theorem 2.5.1. The argument there carries over to show that

$$I^G_{\mathrm{disc},\zeta,c^S}(f) = \operatorname{tr} R^G_{\mathrm{disc},\zeta,c^S}(f), \qquad f \in \mathcal{H}(G(\mathbb{A}_F), K^S),$$

using the fact that a proper Levi subgroup of G is a product of $G'(N)$ with $N' < N$ and a non-quasisplit classical group of lower rank than G; the induction hypothesis is applied to the latter.

Now we consider (2.4.5). Since the stable distributions on the right-hand vanish by Theorem 2.5.1 (if $\mathfrak{e} \in \mathcal{E}^<_{\mathrm{ell}}(G)$, we can also argue as in the proof of that theorem), we deduce that $I^G_{\mathrm{disc},\zeta,c^S}(f) = 0$. Hence $\operatorname{tr} R^G_{\mathrm{disc},\zeta,c^S}$ vanishes as well, and the assertion about $L^2_{\mathrm{disc}}([G])$ follows. □

Theorem 2.5.2 can be rephrased as the existence of a weak endoscopic lift for G as a twisted endoscopic group of $(\tilde{G}(N), \theta)$ in the next corollary. Let us introduce a notion that will be used here and in the next section. Let π_i be a cuspidal automorphic representation of $\mathrm{GL}_{N_i}(\mathbb{A}_F)$ for $i = 1, \ldots, r$. Following [Clozel 1990, Definition 1.2], the isobaric sum of $\pi_1, \dots, \pi_r$, denoted by $\boxplus_{i=1}^r \pi_i$, is defined to be an automorphic representation Π of $\mathrm{GL}_{\sum_i N_i}(\mathbb{A}_F)$ such that Π_v is isomorphic to the Langlands subquotient of the normalized parabolic induction from $\bigotimes_{i=1}^r \pi_{i,v}$ at every place v of F. As remarked in [loc. cit.] an automorphic representation of $\mathrm{GL}_N(\mathbb{A}_F)$ is written as an isobaric sum in a unique way (up to permutation) by a result of Jacquet and Shalika.

Corollary 2.5.3. *Assume (H1). For every discrete automorphic representation π of $G(\mathbb{A}_F)$ unramified away from S, there exists an automorphic representation Π of $G^0(N, \mathbb{A}_F)$, which is an isobaric sum of cuspidal representations, such that $\Pi^\vee \cong \Pi \circ c$ and $(\zeta_{\pi_\infty}, c^S(\pi))$ maps to $(\zeta_{\Pi_\infty}, c^S(\Pi))$ via $\tilde{\xi}$.*

Proof. Since π appears in $L^2_{\mathrm{disc}}([G])^{S-\mathrm{ur}}$, it appears in $L^2_{\mathrm{disc},\zeta,c^S}([G])$ for some (ζ, c^S) mapping to $(\zeta_{\psi,\infty}, c^S(\psi))$ as in Theorem 2.5.2. In particular $(\zeta, c^S) = (\zeta_{\pi_\infty}, c^S(\pi))$. Writing ψ in the form (2.3.1), we can take Π to be the isobaric sum

$$\boxplus_{i\in I}(\mu_i|\det|^{(n_i-1)/2} \boxplus \mu_i|\det|^{(n_i-3)/2} \boxplus \cdots \boxplus \mu_i|\det|^{(1-n_i)/2}).$$

By construction $(\zeta_{\psi,\infty}, c^S(\psi)) = (\zeta_{\Pi_\infty}, c^S(\Pi))$. Since $\psi^\star = \psi$, it follows that $\Pi^\vee \cong \Pi \circ c$. $\square$

3. Automorphic Galois representations

3.1. ***The Buzzard–Gee conjecture.*** Throughout this subsection, let G be a connected reductive group over a number field F (which need not be a classical group). Let ℓ be a prime number and $\iota : \mathbb{C} \xrightarrow{\sim} \overline{\mathbb{Q}}_\ell$ an isomorphism. We work with fixed ℓ and ι at a time, but note that the conjectures below predict the existence of weakly compatible systems of Galois representations in a suitable sense as ℓ and ι vary.

Let $G_{\infty,\mathbb{C}} = \prod_\tau G_\tau$ and $T_{\infty,\mathbb{C}} = \prod_\tau T_\tau$ be as in Section 1.4. Fix a Borel subgroup $B_{\infty,\mathbb{C}} = B_\tau$ containing $T_{\infty,\mathbb{C}}$. The half sum of positive roots is denoted by $\rho_\infty = (\rho_\tau)_\tau \in X^*(T_{\infty,\mathbb{C}})_\mathbb{Q}$. We also view ρ_∞ as the half sum of positive coroots of $\widehat{T}_{\infty,\mathbb{C}}$ relative to $\widehat{B}_{\infty,\mathbb{C}}$, thus an element of $X_*(\widehat{T}_{\infty,\mathbb{C}})_\mathbb{Q}$. We also have $\rho \in X^*(T) = X_*(\widehat{T})$ as the half sum of positive roots for T and B as in Section 1.4. The pairs (B, T) and (B_τ, T_τ) determine isomorphisms $X^*(T) \cong X^*(T_\tau)$ and $X_*(\widehat{T}) \cong X_*(\widehat{T}_\tau)$, under which ρ maps to ρ_τ.

Let $\pi = \otimes'_v \pi_v$ be a discrete automorphic representation of $G(\mathbb{A}_F)$. We assigned the infinitesimal character $\zeta_{\pi_\infty} = (\zeta_{\pi,\tau}) \in X^*(T_{\infty,\mathbb{C}})_\mathbb{C}/\Omega_\infty = \bigoplus_\tau X_*(\widehat{T})_\mathbb{C}/\Omega$ in Section 1.4. We introduce two notions of algebraicity for π in terms of ζ_{π_∞}.

Definition 3.1.1. We say that π is *L-algebraic* if $\zeta_{\pi_\infty} \in X^*(T_{\infty,\mathbb{C}})/\Omega$. If ζ_{π_∞} belongs to the image of $X^*(T_{\infty,\mathbb{C}}) + \rho_\infty$ in $X^*(T_{\infty,\mathbb{C}})_\mathbb{C}/\Omega$ then π is said to be *C-algebraic*. The representation π is *regular* if ζ_{π_∞} is regular as an Ω-orbit in $X^*(T_{\infty,\mathbb{C}})_\mathbb{C}$, i.e., each element of the orbit has the trivial stabilizer in Ω.

The L and C-algebraicity conditions are independent of the choice of $T_{\infty,\mathbb{C}}$ and $B_{\infty,\mathbb{C}}$; see [Buzzard and Gee 2014, Section 2.3]. An equivalent definition can be given by imposing similar conditions on $\zeta_{\pi_{\infty,\tau}}$, T_τ, and ρ_τ for every $\tau : F \hookrightarrow \mathbb{C}$.

Write $S_{\mathrm{ram}}(\pi)$ for the set of places v of F such that either $v \in S_{\mathrm{bad}}(G)$ or π_v is ramified. Let $S(\ell)$ denote the set of places of F above ℓ. At a finite place $v \notin S_{\mathrm{ram}}(\pi)$ of F, let $\phi_{\pi_v} : W_{F_v} \to {}^L G(\mathbb{C})$ denote the unramified L-parameter for π_v (Section 1.4). Changing coefficients by ι, we obtain

$$\iota\phi_{\pi_v} : W_{F_v} \to {}^L G(\overline{\mathbb{Q}}_\ell).$$

Given a Galois representation $r : \Gamma_F \to {}^L G(\overline{\mathbb{Q}}_\ell)$ which is de Rham at ℓ and an embedding $\sigma : F \hookrightarrow \overline{\mathbb{Q}}_\ell$, we follow [Buzzard and Gee 2014, Section 2.4] to assign a Hodge–Tate cocharacter $\mu_{\mathrm{HT}}(r,\sigma) : \mathbb{G}_m \to \iota\widehat{G}$ over $\mathbb{C}_\ell$, whose $\widehat{G}(\mathbb{C}_\ell)$-conjugacy class is defined over $\overline{\mathbb{Q}}_\ell$; here $\iota\widehat{G}$ stands for the base change of $\widehat{G}$ from $\mathbb{C}$ to $\overline{\mathbb{Q}}_\ell$ via ι or its further base extension to $\mathbb{C}_\ell$. (Such a base change is implicit in the notation ${}^L G(\overline{\mathbb{Q}}_\ell)$.) Thereby we obtain a conjugacy class of cocharacters $\mathbb{G}_m \to \iota\widehat{G}$ over $\overline{\mathbb{Q}}_\ell$, which in turn gives an element of $X_*(\iota\widehat{T})/\Omega$. We denote the resulting element by

$$\mu_{r,\sigma} \in X_*(\iota\widehat{T})/\Omega.$$

Conjecture 3.1.2. *Suppose that π is L-algebraic. There exists a Galois representation*

$$r = r_{\ell,\iota}(\pi) : \Gamma_F \to {}^L G(\overline{\mathbb{Q}}_\ell)$$

such that:

(1) $r|^{\mathrm{ss}}_{W_{F_v}} \cong \iota\phi_{\pi_v}$ *at finite places $v \notin S_{\mathrm{ram}}(\pi) \cup S(\ell)$.*

(2) $\mu_{r,\iota\tau} = -\iota\zeta_{\pi,\tau}$ *for every embedding $\tau : F \hookrightarrow \mathbb{C}$.*

Remark 3.1.3. The negative sign in (ii), which does not appear in [Buzzard and Gee 2014, Section 3.2], is due to the different sign convention. (The cyclotomic character has Hodge–Tate weight 1 there; see [loc. cit., Section 2.4].) In this conjecture and the next conjecture, we omit the statement on the image of complex conjugation as we fell short of proving it in the case of interest, see Remark 3.2.8 below.

Remark 3.1.4. When $G = \mathrm{GL}_N$, choosing T to be the diagonal maximal torus, we can identify each member of $X^*(T_{\iota^{-1}\sigma})/\Omega_\tau$ with ordered n integers $(a_i)_{i=1}^n$ with $a_1 \ge a_2 \ge \cdots \ge a_n$. Similarly, each member of $X^*(T_{\infty,\mathbb{C}})_{\mathbb{Q}}/\Omega$ can be regarded as ordered rational numbers $(a_i)_{i=1}^n$ such that $a_1 \ge a_2 \ge \cdots \ge a_n$. In particular, if π is L-algebraic or C-algebraic, then we can write $-\zeta_{\pi,\tau} = (a_i)_{i=1}^n$ for a suitable set of a_i as such. So condition (ii) above may be understood as an equality of multisets for $G = \mathrm{GL}_N$.

Following [Zhu 2020b] (which gives a different but equivalent definition of C-groups as in [Buzzard and Gee 2014]) the C-group of G is defined by taking the semidirect product

$$^{C}G := {}^{L}G \rtimes \mathbb{G}_m, \quad (1 \rtimes t)(g \rtimes 1)(1 \rtimes t)^{-1} = \mathrm{Ad}(\rho(t))g \rtimes 1, \quad g \in {}^{L}G, t \in \mathbb{G}_m.$$

This is well defined because $\mathrm{Ad}(\rho)$ is an algebraic action of $\mathbb{G}_m$ on ${}^{L}G$ (although ρ need not be an algebraic cocharacter into $\hat{G}$). We can also write ${}^{C}G = \widehat{G} \rtimes (\mathbb{G}_m \times \Gamma_F)$ with $\mathbb{G}_m$ and Γ_F acting on $\widehat{G}$ via the $\mathrm{Ad}(\rho)$-action and the L-action respectively, since the Galois action and the $\mathbb{G}_m$-action on $\widehat{G}$ commute. It is convenient to fix a finite Galois extension F'/F over which G splits, and use the finite Galois forms of the L-group ${}^{L}G_{F'/F} = {}^{L}G \rtimes \Gamma_{F'/F}$ and similarly for the C-group ${}^{C}G_{F'/F} = {}^{L}G_{F'/F} \rtimes \mathbb{G}_m$. From now on, we use the finite Galois form and drop F'/F from the subscript unless specified otherwise. We will use the natural $\widehat{G}$-conjugation on ${}^{C}G$, with coefficients in $\overline{\mathbb{Q}}_\ell$ or $\mathbb{C}$, to define the notion of isomorphism for local parameters and global Galois representations valued in ${}^{C}G$. (It does not make any difference if we use the conjugation by $\widehat{G} \rtimes \mathbb{G}_m$ instead.) For the purpose of this section $\mathbb{G}_m$, ${}^{L}G$, ${}^{C}G$, etc. will mean the topological groups of $\overline{\mathbb{Q}}_\ell$ or $\mathbb{C}$-valued points (though they can also be viewed as groups over $\overline{\mathbb{Q}}_\ell$ or $\mathbb{C}$); the coefficient field is suppressed if there is no danger of confusion.

Write $\widehat{T}^{\mathrm{ad}}$ for the image of $\widehat{T}$ in the adjoint group of $\widehat{G}$.

Lemma 3.1.5. *If there exists $\tilde{\rho} \in X_*(\widehat{T})$ which is Γ_F-invariant and has the same image in $X_*(\widehat{T}^{\mathrm{ad}})$ as ρ, then ${}^{C}G \cong {}^{L}G \times \mathbb{G}_m$ via $g \rtimes t \mapsto (g\tilde{\rho}(t), t)$ with the inverse map $(g, t) \mapsto g\tilde{\rho}(t)^{-1} \rtimes t$. These maps are $\widehat{G}$-equivariant: the image of $h(g \rtimes t)h^{-1}$ equals $(hg\tilde{\rho}(t)h^{-1}, t)$ for $h \in \hat{G}$.*

Proof. This is a straightforward verification. □

Let v be a finite place of F not in $S_{\mathrm{ram}}(\pi)$. We introduce a C-normalization of the unramified L-parameter for π_v (with $\mathbb{C}$-coefficient), which is natural from the viewpoint of the geometric Satake equivalence, see [Zhu 2020b, Section 1.4]:

$$\phi^{C}_{\pi_v} : W_{F_v} \to {}^{C}G = {}^{L}G \rtimes \mathbb{G}_m, \quad x \mapsto \phi_{\pi_v}(x) 2\rho(|x|^{1/2}) \rtimes |x|^{-1}. \tag{3.1.1}$$

It is elementary to check that $\phi^{C}_{\pi_v}$ is well defined up to $\widehat{G}$-conjugacy. Indeed, if ϕ_{π_v} is conjugated by an element of $\widehat{G}$ then the resulting $\phi^{C}_{\pi_v}$ is conjugated by the same element. When $\tilde{\rho}$ as in Lemma 3.1.5 exists, the isomorphism therein gives an alternative description of $\phi^{C}_{\pi_v}$:

$$\phi^{C}_{\pi_v} : W_{F_v} \to {}^{L}G \times \mathbb{G}_m, \quad x \mapsto (\phi_{\pi_v}(x) 2(\rho - \tilde{\rho})(|x|^{1/2}), |x|^{-1}). \tag{3.1.2}$$

Example 3.1.6. When G is Sp_{2n} or SO^{η}_{2n}, we take $\tilde{\rho} = \rho$. In this case $F' = F$ except for the case of SO^{η}_{2n} with $\eta \neq 1$; then take $F' = E$. For $G = \mathrm{SO}_{2n+1}$, we take $F' = F$. In this case no $\tilde{\rho}$ as in the lemma exists. For GL_N, we can take $\tilde{\rho} = (N-1, N-2, \ldots, 1, 0)$ with $F' = F$. So when $G = \mathrm{GL}_N$, (3.1.2) reads

$$\phi^C_{\pi_v}(x) = (\phi_{\pi_v}(x)|x|^{(1-N)/2}, |x|^{-1}). \tag{3.1.3}$$

For $G = \mathrm{U}_N$, we take $F' = E$. For odd N we can take $\tilde{\rho} = \rho$, but there does not exist $\tilde{\rho}$ as in Lemma 3.1.5 if N is even. (For instance, $(N-1, N-2, \ldots, 0)$ is not Γ_F-invariant.)

Example 3.1.7. For SO_{2n+1} (with $F' = F$), we have two maps

$$\begin{aligned} &\mathrm{Sp}_{2n} \times \mathbb{G}_m \to \mathrm{GSp}_{2n}, && (g,t) \mapsto gt, \\ &\mathrm{Sp}_{2n} \times \mathbb{G}_m \to {}^C\,\mathrm{SO}_{2n+1} = \mathrm{Sp}_{2n} \rtimes \mathbb{G}_m, && (g,t) \mapsto g2\rho(t)^{-1} \rtimes t^2. \end{aligned}$$

whose kernels are both generated by $(-1,-1)$. This induces an isomorphism

$${}^C\,\mathrm{SO}_{2n+1} \cong \mathrm{GSp}_{2n}\,.$$

Under this isomorphism, (3.1.1) reads

$$\phi^C_{\pi_v} : W_{F_v} \to \mathrm{GSp}_{2n}, \quad x \mapsto \phi_{\pi_v}(x)|x|^{-1/2}.$$

We return to a general discussion. Let $\tau : F \hookrightarrow \overline{\mathbb{Q}}_\ell$ be an embedding. To a Galois representation $r^C : \Gamma_F \to {}^C G(\overline{\mathbb{Q}}_\ell)$ which is de Rham at ℓ, we assign a Hodge–Tate cocharacter $\mu_{\mathrm{HT}}(r^C, \tau) : \mathbb{G}_m \to \widehat{G} \rtimes \mathbb{G}_m$ over $\mathbb{C}_\ell$, which gives rise to an element

$$\mu_{r^C,\tau} \in X_*(\iota\widehat{T} \times \mathbb{G}_m)/\Omega,$$

as in the case of L-group valued representations. Indeed, ${}^C G$ is the L-group of a $\mathbb{G}_m$-extension of G, see [Buzzard and Gee 2014] and [Zhu 2020b], and $\widehat{T} \times \mathbb{G}_m$ is a maximal torus of $\widehat{G} \rtimes \mathbb{G}_m$ whose Weyl group is naturally isomorphic to Ω, the Weyl group for $\widehat{T}$ in $\widehat{G}$. The action of $\omega \in \Omega$ on $X_*(\widehat{T} \times \mathbb{G}_m) = X_*(\widehat{T}) \oplus X_*(\mathbb{G}_m) = X_*(\widehat{T}) \oplus \mathbb{Z}$, induced by the $\widehat{G}$-conjugation on $\widehat{G} \rtimes \mathbb{G}_m$, is that $\omega(a,b) = (\omega a + b(\omega\rho - \rho), b)$, where ωa and $\omega\rho$ are computed using the natural ω-action on $X_*(\widehat{T})$. Define $\zeta^C_{\pi,\tau}$ by

$$-\zeta^C_{\pi,\tau} = (-\zeta_{\pi,\tau} - \rho, 1) \in X_*(\widehat{T} \times \mathbb{G}_m)_{\mathbb{Q}}/\Omega. \tag{3.1.4}$$

This is well defined since if $\zeta_{\pi,\tau} \in X_*(\widehat{T})_{\mathbb{Q}}$ denotes any representative in its Ω-orbit (still denoted $\zeta_{\pi,\tau}$) then $\omega(-\zeta_{\pi,\tau} - \rho, 1) = (-\omega\zeta_{\pi,\tau} - \rho, 1)$ by the preceding formula. When $\tilde{\rho}$ as in Lemma 3.1.5 exists, composition with the isomorphism ${}^C G \cong {}^L G \times \mathbb{G}_m$ gives an alternative description

$$-\zeta^C_{\pi,\tau} = (-\zeta_{\pi,\tau} - \rho + \tilde{\rho}, 1) \in X_*(\widehat{T} \times \mathbb{G}_m)_{\mathbb{Q}}/\Omega. \tag{3.1.5}$$

The reader is cautioned that even though $\widehat{T} \times \mathbb{G}_m$ serves as a maximal torus in both ${}^C G$ and ${}^L G \times \mathbb{G}_m$ via the natural inclusions, the isomorphism ${}^C G \cong {}^L G \times \mathbb{G}_m$

does not induce the identity map on $\widehat{T} \times \mathbb{G}_m$. Rather the induced map "shifts" by $\tilde{\rho}$, which explains the difference between (3.1.4) and (3.1.5). While (3.1.4) is for general ${}^C G$-valued representations, (3.1.5) is for ${}^L G \times \mathbb{G}_m$-valued representations and requires the existence of $\tilde{\rho}$.

The C-algebraic version of Buzzard and Gee's conjecture is adapted to our setting as follows.

Conjecture 3.1.8. *Suppose that π is C-algebraic. There exists a Galois representation*

$$r^C = r^C_{\ell,\iota}(\pi) : \Gamma_F \to {}^C G(\overline{\mathbb{Q}}_\ell)$$

such that:

(1) $r^C|^{\mathrm{ss}}_{W_{F_v}} \cong \iota \phi^C_{\pi_v}$ *at finite places $v \notin S_{\mathrm{ram}}(\pi) \cup S(\ell)$.*

(2) $\mu_{r^C,\iota\tau} = -\iota\zeta^C_{\pi,\tau}$ *for every embedding $\tau : F \hookrightarrow \mathbb{C}$.*

Remark 3.1.9. Condition (i) implies that the composition of r^C with the projection ${}^C G(\overline{\mathbb{Q}}_\ell) \to \mathbb{G}_m(\overline{\mathbb{Q}}_\ell)$ is ω_ℓ^{-1}, the inverse cyclotomic character, in view of (3.1.1). This convention is consistent with [Zhu 2020a] but opposite to that of [Buzzard and Gee 2014, Section 5.3, Conjecture 5.40], where the composition is ω_ℓ.

Remark 3.1.10. When $\rho \in X_*(\widehat{T})$ (not just $\rho \in X_*(\widehat{T})_{\mathbb{Q}}$), Conjectures 3.1.2 and 3.1.8 are equivalent via the isomorphism ${}^C G \cong {}^L G \times \mathbb{G}_m$ of Lemma 3.1.5 given by $\tilde{\rho} = \rho$. Indeed, L-algebraicity coincides with C-algebraicity in that case. Further, r as in the former conjecture gives rise to r^C in the latter conjecture by $r^C(\gamma) := (r(\gamma), \omega_\ell(\gamma)^{-1})$ via the isomorphism. Conversely r can be recovered from r^C by projection.

Conjectures 3.1.2 and 3.1.8 are known for general linear groups under certain hypotheses as we now recall. The case of classical groups will be eventually derived from this result.

Proposition 3.1.11. *Let F, E be as in Section 2.2 and $\star$ as in Section 2.3. Conjectures 3.1.2 and 3.1.8 are true for every discrete automorphic representation π of $\mathrm{GL}_N(\mathbb{A}_E)$* (*in particular E serves as the field F in the conjectures*) *if the following hold*:

- *π is regular* (*and L or C-algebraic as assumed in the conjectures*).
- *$\pi^\star \cong \pi \otimes (\chi \circ N_{E/F})$ for a Hecke character $\chi : F^\times \backslash \mathbb{A}_F^\times \to \mathbb{C}^\times$.*

If π is regular but does not satisfy the second condition, *then Conjectures 3.1.2 and 3.1.8 are true except for the assertions on Hodge–Tate cocharacters.*

Proof. The last assertion will be addressed at the end of proof. Until then we assume that π satisfies both conditions. We begin with the case when π is cuspidal and C-algebraic. Let us represent $\zeta_{\pi,\tau}$ by $(a_1, \dots, a_n) - \left(\frac{1}{2}(n-1), \dots, \frac{1}{2}(n-1)\right)$ with

$(a_i)_{i=1}^n \in \mathbb{Z}^n$. By [Barnet-Lamb et al. 2014, Theorem 2.1.1] (which summarizes a theorem due to many people; the sign condition in that theorem was shown to be superfluous by [Patrikis 2015]), there exists a semisimple Galois representation $R = R_{\ell,\iota}(\pi) : \Gamma_E \to \mathrm{GL}_N(\overline{\mathbb{Q}}_\ell)$ such that

$$R|^{\mathrm{ss}}_{W_{E_v}} \cong \iota\phi_{\pi_v}|\cdot|_v^{(1-N)/2}, \quad v \notin S_{\mathrm{ram}}(\pi) \cup S(\ell), \tag{3.1.6}$$

$$\mu_{R,\iota\tau} = (a_1, \dots, a_n) = -\zeta_{\pi,\tau} + \left(\tfrac{1}{2}(n-1), \dots, \tfrac{1}{2}(n-1)\right). \tag{3.1.7}$$

After choosing $\tilde{\rho}$ as in Example 3.1.6, we identify ${}^C\mathrm{GL}_N \cong \mathrm{GL}_N \times \mathbb{G}_m$ as in Lemma 3.1.5. Then we define an $\mathrm{GL}_N \times \mathbb{G}_m$-valued representation

$$r^C : \Gamma_E \to \mathrm{GL}_N(\overline{\mathbb{Q}}_\ell) \times \mathbb{G}_m(\overline{\mathbb{Q}}_\ell), \quad \gamma \mapsto (R(\gamma), \omega_\ell^{-1}(\gamma)).$$

Comparing (3.1.6) with (3.1.3), we verify part (i) of Conjecture 3.1.8. The cocharacter $\zeta^C_{\pi,\tau}$ in part (ii) of the conjecture becomes a $\mathrm{GL}_N \times \mathbb{G}_m$-valued cocharacter in view of (3.1.5):

$$t \mapsto ((-\zeta_{\pi,\tau} - \rho + \tilde{\rho})(t), t) = \left(\left(-\zeta_{\pi,\tau} + \left(\tfrac{1}{2}(n-1), \dots, \tfrac{1}{2}(n-1)\right)\right)(t), t\right).$$

This coincides with $\mu_{r^C,\iota\tau}$ in view of (3.1.7) and the fact that the Hodge–Tate cocharacter of ω_ℓ^{-1} is the tautological map $t \mapsto t$ on $\mathbb{G}_m$.

We turn to the case of cuspidal L-algebraic π. Then $\pi' := \pi|\det|^{(N-1)/2}$ is cuspidal, regular, and C-algebraic. So there exists $R(\pi')$ such that (3.1.6) and (3.1.7) hold with π' in place of π. We take $r = r_{\ell,\iota}(\pi) := R(\pi')$. Then $r|^{\mathrm{ss}}_{W_{E_v}} \cong \iota\phi_{\pi'_v}|\cdot|_v^{(1-N)/2} \cong \iota\phi_{\pi_v}$ at $v \notin S_{\mathrm{ram}}(\pi) \cup S(\ell)$, so (i) of Conjecture 3.1.2 is satisfied. Similarly (ii) follows from (3.1.7) for $r = R(\pi')$.

From now, let π be a noncuspidal discrete automorphic representation. By [Mœglin and Waldspurger 1989]

$$\pi = \boxplus_{j=1}^r \pi_0|\det|^{(r+1-2j)/2}$$

as an isobaric sum, for some $N_0, r \in \mathbb{Z}_{\geq 1}$ and π_0 a cuspidal automorphic representation of $\mathrm{GL}_{N_0}(\mathbb{A}_E)$, where $N = N_0 r$. If π is regular L-algebraic then $\pi_j := \pi_0|\det|^{(r+1-2j)/2}$ is regular, L-algebraic, and unramified outside $S_{\mathrm{ram}}(\pi)$. By the preceding argument, we have $r_{\ell,\iota}(\pi_j)$ corresponding to π_j satisfying Conjecture 3.1.2. Then $r := \bigoplus_j r_{\ell,\iota}(\pi_j)$ is the Galois representation corresponding to π predicted by the conjecture. We leave to the reader to verify Conjecture 3.1.8 when π is regular C-algebraic and noncuspidal as no new idea is needed.

Finally, if the second condition on π is not assumed, we can run the same argument as above except that we apply the theorems of Harris, Lan, Taylor and Thorne [2016] and Scholze [2015] instead of [Barnet-Lamb et al. 2014, Theorem 2.1.1] to obtain Galois representations. The only difference in the outcome is that the Hodge–Tate weights have not been identified for the Galois representations in [Harris et al. 2016; Scholze 2015], so we are unable to verify (ii) in Conjectures 3.1.2 and 3.1.8. □

3.2. ***Existence of Galois representations for classical groups.*** From here until the end of the paper, we use the same notation as in Section 2.5, including G^*, N, and $\tilde{\xi} : {}^L G \hookrightarrow {}^L\tilde{G}^0(N)$. In Case U, take $\tilde{\xi}$ to be the standard base change morphism $\tilde{\xi}_+$ (rather than $\tilde{\xi}_-$). In Case S, we recall that $\tilde{\xi}|_{\widehat{G}^*}$ is the standard embedding of $\widehat{G}^*$ into GL_N.

Let ℓ be a prime and choose an isomorphism $\iota : \mathbb{C} \simeq \overline{\mathbb{Q}}_\ell$. Let S be a finite set of places of F which contains all places above ℓ and ∞ such that G is unramified at places outside S.

Definition 3.2.1. A discrete automorphic representation π of $G(\mathbb{A}_F)$ is said to be *std-regular* if $\tilde{\xi}(\zeta_{\pi_\infty}) \in \mathcal{C}_\infty(N)$ is regular.

Lemma 3.2.2. *If π is std-regular then it is regular. The two conditions are equivalent unless G is an inner form of* SO^η_{2n}.

Proof. As we explicated the map $X_*(\widehat{T}) \to X_*(\widehat{T}_N)$ induced by $\tilde{\xi}$ in Section 2.2, the lemma follows from the definition. □

Example 3.2.3. When $G = \mathrm{SO}^\eta_{2n}$, a Weyl group orbit in $X_*(\widehat{T}) = \mathbb{Z}^n$ is uniquely represented by (a_i) such that $a_1 \geq a_2 \geq \cdots \geq a_{n-1} \geq |a_n|$. If ζ_{π_∞} corresponds to such a tuple (a_i) then π is regular if strict inequalities hold everywhere, and std-regular if furthermore $a_n \neq 0$.

Let $r : \Gamma_E \to \mathrm{GL}_m(\overline{\mathbb{Q}}_\ell)$ be a Galois representation. Define another representation $r^\perp$ by

$$r^\perp(\gamma) := {}^t r(c\gamma c^{-1})^{-1},$$

which is isomorphic to the dual representation $r^\vee$ in Case S. Let $\chi : \Gamma_E \to \overline{\mathbb{Q}}_\ell^\times$ be a Galois character such that $\chi(c\gamma c^{-1}) = \chi(\gamma)$ for all $\gamma \in \Gamma_E$ (which is automatic in Case S). From now assume that r is irreducible. Provided that $r^\perp \cong r\chi$, we recall how to define a sign

$$\mathrm{sgn}(r, \chi) \in \{\pm 1\}$$

following [Bellaïche and Chenevier 2011, Section 1.1]. In Case S, we obtain a nonzero Γ_F-equivariant pairing $r \otimes r \to \chi^{-1}$ up to a nonzero scalar. According to whether the pairing is orthogonal or symplectic (it cannot be both since r is irreducible), we assign 1 or -1 as the value of $\mathrm{sgn}(r, \chi)$. When χ is trivial, we just write $\mathrm{sgn}(r)$ and refer to it as the sign of r. Of course if m is odd then always $\mathrm{sgn}(r, \chi) = 1$. In Case U, by assumption there exists $h \in \mathrm{GL}_m(\overline{\mathbb{Q}}_\ell)$, unique up to nonzero scalars, such that $r^\perp = hrh^{-1}\chi$. Then it is elementary to check that ${}^t h = \mathrm{sgn}(r, \chi)h$ for $\mathrm{sgn}(r, \chi) \in \{\pm 1\}$, which does not depend on the choice of h.

Henceforth we restrict E and F as follows in order to access Proposition 3.1.11:

(Case S) $E = F$ is a totally real field.

(Case U) F is a totally real field, and E is a CM quadratic extension of F.

Consider the following hypotheses — see the paragraph above Theorem 1.2.2. The two versions of (H2) are equivalent to each other since $\zeta_{\Pi_\infty} = \tilde{\xi}(\zeta_{\pi_\infty})$.

(H2) π is std-regular.

(H3) In Corollary 2.5.3, if Π is written as an isobaric sum $\Pi = \boxplus_{i=1}^{r} \Pi_i$ then Π_i is (conjugate) self-dual for every i, i.e., $\Pi_i^\star = \Pi_i$.

Proposition 3.2.4. *Let E and F be as above. Assume (H1). Let π be a discrete automorphic representation of $G(\mathbb{A}_F)$ which is unramified outside S, C-algebraic, and satisfying (H2) and (H3). Then there exists a continuous semisimple Galois representation*

$$R = R_{\ell,\iota}(\pi) : \Gamma_{E,S} \to \mathrm{GL}_N(\overline{\mathbb{Q}}_\ell)$$

with the following property. If $G^* = \mathrm{Sp}_{2n}$ or SO_{2n}^{η} (Case S), we have:

(i) $R|_{W_{F_v}}^{\mathrm{ss}} \cong \iota\tilde{\xi}\phi_{\pi_v}$ *for every place v of F not above S.*

(ii) $\mu_{R,\iota\sigma} = -\iota\xi(\zeta_{\pi,\sigma})$ *for embeddings $\sigma : F \hookrightarrow \mathbb{C}$.*

(iii) $R^\vee \cong R$. *When $G^* = \mathrm{SO}_{2n}^{\eta}$, every self-dual irreducible constituent of R has sign* 1.

(iv) $\det R = \mathbf{1}$ *if $G^* = \mathrm{Sp}_{2n}$ and* $\det R = \eta$ *if $G^* = \mathrm{SO}_{2n}^{\eta}$.*

If $G^* = \mathrm{SO}_{2n+1}$ (Case S) then:

(i') $R|_{W_{F_v}}^{\mathrm{ss}} \cong \iota(\tilde{\xi}\phi_{\pi_v}|\cdot|^{(1-N)/2})$ *for every place v of F not above S.*

(ii') $\mu_{R,\iota\sigma} = -\iota\tilde{\xi}(\zeta_{\pi,\sigma}) + \big(\frac{1}{2}(N-1), \dots, \frac{1}{2}(N-1)\big)$ *for embeddings $\sigma : F \hookrightarrow \mathbb{C}$.*

(iii') $R^\perp \cong R \otimes \omega_\ell^{N-1}$. *For every irreducible constituent r of R such that $r^\perp \cong r \otimes \omega_\ell^{N-1}$, we have* $\mathrm{sgn}(r, \omega_\ell^{N-1}) = -1$.

If $G^* = \mathrm{U}_N$ (Case U) then with $\tilde{\xi}_u, \tilde{\xi}_\tau$ as in Section 2.1:

(i'') $R|_{W_{E_u}}^{\mathrm{ss}} \cong \iota(\tilde{\xi}_u\phi_{\pi_v}|\cdot|_u^{(1-N)/2})$ *for every place u of E not above S, where v is the place of F restricted from u.*

(ii'') $\mu_{R,\iota\tau} = -\iota\tilde{\xi}_\tau(\zeta_{\pi,\tau|_F}) + (\frac{N-1}{2}, \dots, \frac{N-1}{2})$ *for embeddings $\tau : E \hookrightarrow \mathbb{C}$.*

(iii'') $R^\perp \cong R \otimes \omega_\ell^{N-1}$. *For every irreducible constituent r of R such that $r^\perp \cong r \otimes \omega_\ell^{N-1}$, we have* $\mathrm{sgn}(r, \omega_\ell^{N-1}) = 1$.

If (H1) and (H2) are assumed but not (H3), then the above is true except (ii), (ii'), and (ii'').

Remark 3.2.5. In fact the proof below shows that every irreducible constituent of R in (iii) (resp. (iii') and (iii")) is self-dual (resp. self-dual up to ω_ℓ^{N-1}) thanks to (H3).

Remark 3.2.6. We could have stated the U_N-case uniformly with the SO_{2n+1}-case if we rewrite R as a Galois representation $\Gamma_{F,S} \to {}^L G(N)(\overline{\mathbb{Q}}_\ell)$ via a variant of Shapiro's lemma. Then (i") and (ii") can be merged into (i') and (ii'). E.g., both (i') and (i") assert $R|^{\mathrm{ss}}_{W_{F_v}} \cong \iota\tilde{\xi}\phi_{\pi_v}|\cdot|_v^{(1-N)/2}$ in this formulation. However the current formulation for unitary groups is convenient in Section 3.4.

Proof. Let $\Pi = \boxplus_{i=1}^r \Pi_i$ be the automorphic representation of $G(N, \mathbb{A}_F) = \mathrm{GL}_N(\mathbb{A}_E)$ which is a functorial lift of π as in Corollary 2.5.3. We are going to apply Proposition 3.1.11 to each Π_i. The proof will be presented only when (H1), (H2), and (H3) are assumed. If (H3) is dropped then we lose track of Hodge–Tate cocharacters according to Proposition 3.1.11 but the argument is identical other than that. This explains the last assertion of Proposition 3.2.4.

According to (H3), each Π_i is a cuspidal automorphic representation of $\mathrm{GL}_{m_i}(\mathbb{A}_E)$ such that $\Pi_i^\star \cong \Pi_i$ and $\sum_i m_i = N$. Since $(\zeta_{\Pi_\infty}, c^S(\Pi)) = \tilde{\xi}(\zeta_{\pi_\infty}, c^S(\pi))$, the std-regularity of π implies that Π is regular. Moreover the description of ρ and $\tilde{\xi}$ in Sections 2.1 and 2.2 tells us that:

- If $G^* = \mathrm{Sp}_{2n}$ then π is also L-algebraic; Π is both L and C-algebraic.
- If $G^* = \mathrm{SO}_{2n}^\eta$ then π is also L-algebraic; Π is L-algebraic but not C-algebraic.
- If $G^* = \mathrm{SO}_{2n+1}$ then Π is C-algebraic but not L-algebraic.
- If $G^* = \mathrm{U}_N$ then Π is C-algebraic; it is not L-algebraic if N is even.

Suppose $G^* = \mathrm{SO}_{2n+1}$. Since Π is regular C-algebraic, we see that $\Pi|\det|^{(1-N)/2}$ is regular L-algebraic, so $\Pi_i' := \Pi_i|\det|^{(1-N)/2}$ is regular L-algebraic as well. Moreover $(\Pi_i')^\star \cong \Pi_i'|\det|^{N-1}$, so Proposition 3.1.11 yields a Galois representation $r_i' := r_{\ell,\iota}(\Pi_i')$. Then $R := \bigoplus_{i=1}^r r_i'$ satisfies (i') and (ii') in light of properties (i) and (ii) of Conjecture 3.1.2 for r_i'. Indeed, (i') is checked as follows:

$$R|^{\mathrm{ss}}_{W_{F_v}} \cong \iota\phi_{\Pi_v'} \cong \iota\phi_{\Pi_v}|\cdot|^{(1-N)/2} \cong \iota\tilde{\xi}\phi_{\pi_v}|\cdot|^{(1-N)/2}, \quad v \notin S.$$

As for (ii'), since $\mu_{r_i',\iota\sigma} = \iota\zeta_{\Pi_i',\sigma}$ for every i, we have

$$\mu_{R,\iota\sigma} = -\iota\zeta_{\Pi|\det|^{(1-N)/2},\sigma} = -\iota\tilde{\xi}(\zeta_{\pi,\sigma}) + \big(\tfrac{1}{2}(N-1), \dots, \tfrac{1}{2}(N-1)\big).$$

Moreover, we have $\phi_{\Pi_v}^\vee \cong \phi_{\Pi_v}$ since $\Pi^\vee \cong \Pi$, so the displayed formula implies that $R^\perp \cong R \otimes \omega_\ell^{N-1}$. The rest of (iii') is verified by [Bellaïche and Chenevier 2011, Corollary 1.3] (their n is our N, which is even; their η_λ is trivial). This finishes the proof when G^* is SO_{2n+1}.

The case $G^* = \mathrm{U}_N$ can be treated as in the SO_{2n+1}-case, by defining Π'_i, r'_i, and R in the same way. There is only a minor difference in showing (i"):

$$R|^{\mathrm{ss}}_{W_{E_u}} \cong \iota\phi_{\Pi'_u} \cong \iota\phi_{\Pi_u}|\cdot|^{(1-N)/2} \cong \iota\tilde{\xi}_u\phi_{\pi_v}|\cdot|^{(1-N)/2}, \quad v \notin S.$$

The justification of (ii') also goes through for (ii") with a similar change. The proof of (iii") is identical to that of (iii') except that we use the conjugate duality and invoke [Bellaïche and Chenevier 2011, Theorem 1.2] rather than Corollary 1.3 therein.

Now consider $G^* = \mathrm{Sp}_{2n}$ or SO^{η}_{2n}. Then Π is regular L-algebraic so each Π_i is regular L-algebraic, cuspidal, and $\Pi_i^{\vee} \cong \Pi_i$. By Proposition 3.1.11, there is a corresponding Galois representation $r_i := r_{\ell,\iota}(\Pi_i)$. Taking $R := \bigoplus_{i=1}^r r_i$, we deduce (i) and (ii) for R from the properties of r_i as in the preceding paragraph. It follows from (i) that R is self-dual. When $G^* = \mathrm{SO}^{\eta}_{2n}$, [Bellaïche and Chenevier 2011, Corollary 1.3] (their n is our N, which is even; their η_λ equals our ω_ℓ^{1-N} in the case at hand, so $\eta_\lambda(c) = -1$) tells us that the irreducible self-dual constituents of R are orthogonal, so the proof of (iii) is complete. Finally, to show (iv), it suffices to check that $\det R|_{W_{F_v}}$ equals $\mathbf{1}$ if $G^* = \mathrm{Sp}_{2n}$ and η_v if $G^* = \mathrm{SO}^{\eta}_{2n}$ for $v \notin S$. This follows from part (i). Indeed, this is obvious if $G^* = \mathrm{Sp}_{2n}$ since the image of $\tilde{\xi}$ is contained in SO_{2n+1}; if $G^* = \mathrm{SO}^{\eta}_{2n}$, it is enough to note that the composite map $\det \circ \tilde{\xi} : {}^L\mathrm{SO}^{\eta}_{2n} \to \mathrm{GL}_{2n} \to \mathbb{G}_m$ is given by the projection ${}^L\mathrm{SO}^{\eta}_{2n} \twoheadrightarrow \mathrm{Gal}(F_\eta/F)$ followed by η. □

When $\phi_1, \phi_2 : W_{F_v} \to {}^C G(\overline{\mathbb{Q}}_\ell)$ are two parameters, we write $\phi_1 \overset{\circ}{\cong} \phi_2$ to mean

- $\phi_1 \cong \phi_2$ if $G^* \ncong \mathrm{SO}^{\eta}_{2n}$, and
- $\phi_1 \cong \phi_2$ or $\hat{\theta}^\circ(\phi_1) \cong \phi_2$ if $G^* \cong \mathrm{SO}^{\eta}_{2n}$.

Similarly if $\mu_1, \mu_2 \in X_*(\widehat{T})_{\mathbb{Q}}/\Omega$ then $\mu_1 \overset{\circ}{=} \mu_2$ means $\mu_1 = \mu_2$ if $G^* \ncong \mathrm{SO}^{\eta}_{2n}$, and $\mu_1 = \mu_2$ or $\hat{\theta}^\circ(\mu_1) = \mu_2$ if $G^* \ncong \mathrm{SO}^{\eta}_{2n}$.

Theorem 3.2.7. *Let E and F be as above and assume (H1). Let π be as in Proposition 3.2.4 satisfying (H2) and (H3). Then Conjecture 3.1.8 holds true if $G^* \ncong \mathrm{SO}^{\eta}_{2n}$, and it holds up to outer automorphism if $G^* \cong \mathrm{SO}^{\eta}_{2n}$. More precisely, there exists a continuous semisimple Galois representation*

$$r^C = r^C_{\ell,\iota}(\pi) : \Gamma_{F,S} \to {}^C G(\overline{\mathbb{Q}}_\ell)$$

such that:

(1) $r^C|^{\mathrm{ss}}_{W_{F_v}} \overset{\circ}{\cong} \iota\phi^C_{\pi_v}$ *for every place v of F not above S.*

(2) $\mu_{r^C,\iota\sigma} \overset{\circ}{=} -\iota\zeta^C_{\pi,\sigma}$ *for every $\sigma : F \hookrightarrow \overline{\mathbb{Q}}_\ell$.*

If we drop (H3), then the theorem still holds true except for part (ii).

The proof is the same whether we assume (H3) or not. Without (H3), we lose property (ii) of the theorem only because we do not know (ii), (ii'), and (ii") in Proposition 3.2.4. With this understanding, we will present the proof in Section 3.3 and Section 3.4 below in the case that all of (H1), (H2), and (H3) are assumed.

Remark 3.2.8. Buzzard and Gee also makes a prediction on the image of complex conjugation at each real place but we do not see how to prove it completely beyond some partial results. For instance, in the proof of Proposition 3.2.4 in Case S, every r'_i is totally odd by [Taylor 2012; Taïbi 2016; Caraiani and Le Hung 2016], but this alone does not determine the image of complex conjugation (up to conjugacy) under R. Thus the information is insufficient to pin down the image of complex conjugation under r^C in Theorem 3.2.7. The image is sometimes identified under additional hypotheses; see [Kret and Shin 2020, Theorem 6.5; 2023, Theorem 2.4].

3.3. ***Proof of Theorem 3.2.7: Case S.*** Write $R = R_{\ell,\iota}(\pi) : \Gamma_F \to \mathrm{GL}_N(\overline{\mathbb{Q}}_\ell)$ for the Galois representation as in Proposition 3.2.4. (We are in the $E = F$ case.) We will divide into three cases according to G^*. When G^* is either Sp_{2n} or SO^η_{2n}, we will prove Conjecture 3.1.2 as this is equivalent to Theorem 3.2.7 but notationally simpler; see Remark 3.1.10.

If $G^* = \mathrm{Sp}_{2n}$ then $R^\vee \cong R$ and every self-dual irreducible constituent is orthogonal by (iii) of Proposition 3.2.4. Hence, possibly after a GL_{2n+1}-conjugation, R factors as

$$\Gamma_{E,S} \longrightarrow \mathrm{O}_{2n+1}(\overline{\mathbb{Q}}_\ell) \xrightarrow{\tilde{\xi}} \mathrm{GL}_{2n+1}(\overline{\mathbb{Q}}_\ell).$$

Take $r^C_{\ell,\iota}(\pi) : \Gamma_{E,S} \to \mathrm{O}_{2n+1}(\overline{\mathbb{Q}}_\ell)$ to be the first map. By Proposition 3.2.4(iv), the image of $r^C_{\ell,\iota}(\pi)$ is contained in $\mathrm{SO}_{2n+1}(\overline{\mathbb{Q}}_\ell)$. Since the natural map $\widehat{T}/\Omega \to \widehat{T}_{2n+1}/\Omega_{2n+1}$ is injective, one deduces (i) and (ii) of Conjecture 3.1.2 from (i) and (ii) of Proposition 3.2.4.

Next consider $G^* = \mathrm{SO}^\eta_{2n}$. As in the Sp_{2n}-case, again from Proposition 3.2.4(iii), we obtain

$$r^C_{\ell,\iota}(\pi) : \Gamma_{F,S} \to \mathrm{O}_{2n}(\overline{\mathbb{Q}}_\ell)$$

such that $\iota(\eta) \circ r^C_{\ell,\iota}(\pi) \cong R_{\ell,\iota}(\pi)$. The difference is that $\widehat{T}/\Omega \to \widehat{T}_{2n}/\Omega_{2n}$ is not a bijection but induces a bijection on the set of $\hat{\theta}^\circ$-orbits on $\widehat{T}/\Omega \to \widehat{T}_{2n}$ onto $\widehat{T}_{2n}/\Omega_{2n}$. With this observation, (i) and (ii) of Conjecture 3.1.2 are implied by (i) and (ii) of Proposition 3.2.4.

In the remaining case $G^* = \mathrm{SO}_{2n+1}$, we identify ${}^C\mathrm{SO}_{2n+1} = \mathrm{GSp}_{2n}$ as in Example 3.1.7. Let $R = R_{\ell,\iota}(\pi) : \Gamma_F \to \mathrm{GL}_{2n}(\overline{\mathbb{Q}}_\ell)$ be the Galois representation corresponding to π by Proposition 3.2.4. By (iii') of the proposition, there is a symplectic pairing $(R \otimes \omega_\ell^{n-1}) \otimes (R \otimes \omega_\ell^{n-1}) \to \omega_\ell^{-1}$. After conjugation, $R \otimes \omega_\ell^{n-1}$ factors through the standard embedding $\tilde{\eta}^C : \mathrm{GSp}_{2n} \to \mathrm{GL}_{2n}$. Denote the resulting representation by

$$r^C = r^C_{\ell,\iota}(\pi) : \Gamma_{F,S} \to \mathrm{GSp}_{2n}(\overline{\mathbb{Q}}_\ell).$$

Write $\lambda : \mathrm{GSp}_{2n} \to \mathbb{G}_m$ for the similitude character. Since the symplectic pairing is valued in ω_ℓ^{-1}, we have

$$\lambda r^C = \omega_\ell^{-1}.$$

By construction, the properties of R in Proposition 3.2.4 tell us that

$$\tilde{\eta}^C(r^C|^{\mathrm{ss}}_{W_{F_v}}) \cong \iota(\tilde{\eta}\phi_{\pi_v}\cdot|\cdot|^{-1/2}) = \tilde{\eta}^C(\iota\phi_{\pi_v}\cdot|\cdot|^{-1/2}),$$
$$\tilde{\eta}^C(\mu_{r^C,\iota\sigma}) = \mu_{\tilde{\eta}^C r^C,\iota\sigma} = -\iota\tilde{\eta}(\zeta_{\pi,\sigma}) + \left(\tfrac{1}{2},\ldots,\tfrac{1}{2}\right) = \tilde{\eta}^C\left(-\iota\zeta_{\pi,\sigma} + \left(\tfrac{1}{2},\ldots,\tfrac{1}{2}\right)\right).$$

On the other hand, we have

$$\lambda(\iota(\tilde{\eta}\phi_{\pi_v}\cdot|\cdot|^{-1/2})) = |\cdot|^{-1} = \lambda r^C|_{W_{F_v}} = \lambda(r^C|^{\mathrm{ss}}_{W_{F_v}}),$$
$$\lambda\left(-\iota\zeta_{\pi,\sigma} + \left(\tfrac{1}{2},\ldots,\tfrac{1}{2}\right)\right) = 1 = \mu_{\omega_\ell^{-1},\iota\sigma} = \mu_{\lambda r^C,\iota\sigma} = \lambda(\mu_{r^C,\iota\sigma}).$$

To deduce the theorem, we need to show that the above relations hold without taking $\tilde{\eta}^C$ and λ at both ends. This is implied by the following facts. Firstly, if semisimple elements $g_1, g_2 \in \mathrm{GSp}_{2n}(\overline{\mathbb{Q}}_\ell)$ are such that $\tilde{\eta}^C(g_1)$, $\tilde{\eta}^C(g_2)$ are conjugate and $\lambda(g_1) = \lambda(g_2)$ then g_1, g_2 are conjugate in $\mathrm{GSp}_{2n}(\overline{\mathbb{Q}}_\ell)$; see [Kret and Shin 2023, Lemmas 1.1, 1.3]. Secondly, the analogous injectivity is also true on the level of conjugacy classes of cocharacters via the isomorphism $X_*(T_{\mathrm{GSp}})\otimes_{\mathbb{Z}}\overline{\mathbb{Q}}_\ell^\times \cong T_{\mathrm{GSp}}(\overline{\mathbb{Q}}_\ell)$, which is equivariant for the Weyl group action, where T_{GSp} is a maximal torus of GSp_{2n} over $\overline{\mathbb{Q}}_\ell$. The proof in the SO_{2n+1}-case is complete.

3.4. ***Proof of Theorem 3.2.7: Case U.*** Recall that E is a CM quadratic extension of a totally real field F in this case. Throughout this section we set

$$\tilde{\rho}(t) := \mathrm{diag}(t^{N-1}, t^{N-2}, \ldots, t, 1) \in \mathrm{GL}_N(\overline{\mathbb{Q}}_\ell) \cong \hat{\mathrm{U}}_N(\overline{\mathbb{Q}}_\ell), \quad t \in \mathbb{G}_m,$$

where the isomorphism is fixed as in Section 2.1. (The same $\tilde{\rho}$ appeared in Example 3.1.6 for odd unitary groups. Here $\tilde{\rho}$ is also considered for even unitary groups as Lemma 3.1.5 is irrelevant here.) A key point in the proof is to extend a GL_N-valued representation of $\Gamma_{E,S}$ to a ${}^C\mathrm{U}$-valued representation of $\Gamma_{F,S}$. We begin with two lemmas to help address this problem. Similar problems were considered in related settings; see [Clozel et al. 2008, Section 2.1; Bellaïche and Chenevier 2009, Appendix A.11; Barnet-Lamb et al. 2014, Section 1] (see [Buzzard and Gee 2014, Section 8.3] for a comparison with C-groups), and [Kret and Shin 2020, Appendix A] for instance.

Lemma 3.4.1. *Let $R : \Gamma_{E,S} \to \mathrm{GL}_N(\overline{\mathbb{Q}}_\ell)$ be a Galois representation. If there exists $h \in \mathrm{GL}_N(\overline{\mathbb{Q}}_\ell)$ such that*

$$ {}^t h = h \quad \text{and} \quad R^\perp(\gamma) = hR(\gamma)h^{-1} \cdot \omega_\ell(\gamma)^{N-1}, \quad \gamma \in \Gamma_{E,S}, \tag{3.4.1}$$

then there exists a Galois representation

$$\tilde{R} : \Gamma_{F,S} \to {}^C\mathrm{U}_N(\overline{\mathbb{Q}}_\ell) = \mathrm{GL}_N(\overline{\mathbb{Q}}_\ell) \rtimes (\mathbb{G}_m \times \{1, c\})$$

uniquely determined by:

- $\tilde{R}(\gamma) = R(\gamma)\tilde{\rho}(\omega_\ell(\gamma)) \rtimes (\omega_\ell^{-1}(\gamma), 1)$ *for all* $\gamma \in \Gamma_{E,S}$.
- $\tilde{R}(c) = h^{-1}J_N \rtimes (-1, c)$.

Proof. The uniqueness is clear. The main point is to check that the two conditions on $\tilde{R}$ define a group homomorphism. This amounts to checking that $\tilde{R}(c)^2 = 1$ and $\tilde{R}(c)\tilde{R}(\gamma)\tilde{R}(c)^{-1} = \tilde{R}(c\gamma c^{-1})$ for $\gamma \in \Gamma_{E,S}$. Set $h_0 := h^{-1}J_N = h^{-1}J_N^{-1}$ and let $\tilde{\rho}$ be as in Example 3.1.6. We compute

$$\begin{aligned}
\tilde{R}(c)^2 &= (h_0 \rtimes (-1, c))(h_0 \rtimes (-1, c)) = (h_0 \rtimes (-1, 1))(J_N^{*}{}^t h_0^{-1} J_N^{*,-1} \rtimes (-1, 1)) \\
&= h_0\tilde{\rho}(-1)J_N^{*}{}^t h_0^{-1} J_N^{*,-1}\tilde{\rho}(-1)^{-1} = h_0 J_N {}^t h_0^{-1} J_N^{-1} = h^{-1}{}^t h = 1.
\end{aligned}$$

$$\begin{aligned}
&\tilde{R}(c)\tilde{R}(\gamma)\tilde{R}(c)^{-1} \\
&\quad = (h_0 \rtimes (-1, c))(R(\gamma)\tilde{\rho}(\omega_\ell(\gamma)) \rtimes (\omega_\ell^{-1}(\gamma), 1))(h_0 \rtimes (-1, c))^{-1} \\
&\quad = (h_0 \rtimes (-1, 1))(J_N^{*}{}^t R(\gamma)^{-1}\tilde{\rho}(\omega_\ell(\gamma))^{-1} J_N^{*,-1} \\
&\qquad\qquad \rtimes (\omega_\ell^{-1}(\gamma), 1))(h_0 \rtimes (-1, 1))^{-1} \\
&\quad = h_0 J_N ({}^t R(\gamma)^{-1}\tilde{\rho}(\omega_\ell(\gamma))^{-1} J_N^{-1} \rtimes (\omega_\ell^{-1}(\gamma), 1))h_0^{-1} \\
&\quad = h^{-1}{}^t R(\gamma)^{-1}\tilde{\rho}(\omega_\ell(\gamma))^{-1} J_N^{-1}\tilde{\rho}(\omega_\ell(\gamma))^{-1} h_0^{-1}\tilde{\rho}(\omega_\ell(\gamma)) \rtimes (\omega_\ell^{-1}(\gamma), 1).
\end{aligned}$$

By an explicit computation with $\tilde{\rho}$ and J_N, we verify that

$$J_N^{-1}\tilde{\rho}(\omega_\ell(\gamma))^{-1} = \tilde{\rho}(\omega_\ell(\gamma))J_N^{-1}\omega_\ell(\gamma)^{1-N}.$$

Substituting in the above formula and using $h = J_N^{-1}h_0^{-1}$, we obtain

$$\tilde{R}(c)\tilde{R}(\gamma)\tilde{R}(c)^{-1} = h^{-1} \cdot {}^t R(\gamma)^{-1} h\tilde{\rho}(\omega(\gamma)) \cdot \omega_\ell(\gamma)^{1-N} \rtimes (\omega_\ell(\gamma)^{-1}, 1).$$

On the other hand, we see from (3.4.1) that

$$R(c\gamma c^{-1}) = {}^t R^\perp(\gamma)^{-1} = h^{-1}{}^t R(\gamma)^{-1} h \cdot \omega_\ell(\gamma)^{1-N}$$

so $\tilde{R}(c\gamma c^{-1}) = h^{-1}{}^t R(\gamma)^{-1} h \cdot \omega_\ell(\gamma)^{1-N}\tilde{\rho}(\omega_\ell(\gamma)) \rtimes (\omega_\ell^{-1}(\gamma), 1)$. We conclude that $\tilde{R}(c)\tilde{R}(\gamma)\tilde{R}(c)^{-1} = \tilde{R}(c\gamma c^{-1})$, recalling that $\omega_\ell(\gamma)$ lies in the center of $\mathrm{GL}_N(\overline{\mathbb{Q}}_\ell)$. □

Lemma 3.4.2. *Let $R : \Gamma_{E,S} \to \mathrm{GL}_N(\overline{\mathbb{Q}}_\ell)$ be a semisimple Galois representation such that*:

- $R^\star \cong R \otimes \omega_\ell^{N-1}$.
- *Every irreducible subrepresentation $R_0 \subset R$ such that $R_0^\star \cong R_0 \otimes \omega_\ell^{N-1}$ has* $\mathrm{sgn}(R_0, \omega_\ell^{N-1}) = 1$.

Then there exists a Galois representation

$$\tilde{R} : \Gamma_{F,S} \to {}^C\,\mathrm{U}_N(\overline{\mathbb{Q}}_\ell) = \mathrm{GL}_N(\overline{\mathbb{Q}}_\ell) \rtimes (\mathbb{G}_m \times \{1, c\})$$

such that:

- $\tilde{R}(\gamma) = R(\gamma)\tilde{\rho}(\omega_\ell(\gamma)) \rtimes (\omega_\ell^{-1}(\gamma), 1)$ *for all* $\gamma \in \Gamma_{E,S}$.
- $\tilde{R}(c) = h^{-1} J_N \rtimes (-1, c)$ *for a symmetric matrix* $h \in \mathrm{GL}_N(\overline{\mathbb{Q}}_\ell)$.

Proof. Since $R^\star \cong R \otimes \omega_\ell^{N-1}$, we can decompose R into irreducibles

$$R \cong \left(\bigoplus_{i=1}^{r} R_i \right) \oplus \left(\bigoplus_{j=1}^{s} (R_j \oplus (R_j^\perp \otimes \omega_\ell^{1-N})) \right)$$

such that $R_i^\star \cong R_i \otimes \omega_\ell^{N-1}$ and $R_j^\star \not\cong R_j \otimes \omega_\ell^{N-1}$ for every i, j. (Recall that $R_j^\star \cong R_j^\perp$.) Write $d_i := \dim R_i$ and $d_j := \dim R_j$. For each i, since $\mathrm{sgn}(R_i, \omega_\ell^{N-1}) = 1$, there exists $h_i \in \mathrm{GL}_{d_i}(\overline{\mathbb{Q}}_\ell)$ satisfying (3.4.1) for h_i and R_i in place of h and R. For $1 \le j \le s$, take

$$h_j := \begin{pmatrix} 0 & I \\ I & 0 \end{pmatrix} \in \mathrm{GL}_{2d_j}(\overline{\mathbb{Q}}_\ell),$$

where 0 and I stand for the zero and identity $d_j \times d_j$ matrices. Then it satisfies (3.4.1) for h_j and $R_j^\perp \otimes \omega_\ell^{1-N}$ in place of h and R by construction. Hence if we form $h \in \mathrm{GL}_N(\overline{\mathbb{Q}}_\ell)$ as a block diagonal matrix according to the decomposition of R by putting together h_i and h_j, then (3.4.1) holds true for h and R. By Lemma 3.4.1 we obtain the desired $\tilde{R}$. □

Now we put ourselves in the setting of Theorem 3.2.7 for $G^* = \mathrm{U}_N$ and let $R : \Gamma_{E,S} \to \mathrm{GL}_N(\overline{\mathbb{Q}}_\ell)$ be the representation coming from Proposition 3.2.4. Since R satisfies the condition of Lemma 3.4.2, we obtain

$$r^C : \Gamma_{F,S} \to {}^C\,\mathrm{U}_N(\overline{\mathbb{Q}}_\ell) \overset{\tilde{\xi}_0}{=} \mathrm{GL}_N(\overline{\mathbb{Q}}_\ell) \rtimes (\{1, c\} \times \mathbb{G}_m)$$

as in the lemma. (We renamed $\tilde{R}$ as r^C.) By construction the following composition is equal to the representation (R, ω_ℓ^{-1}):

$$\Gamma_{E,S} \xrightarrow{r^C} \mathrm{GL}_N(\overline{\mathbb{Q}}_\ell) \rtimes \mathbb{G}_m \xrightarrow{\varsigma} \mathrm{GL}_N(\overline{\mathbb{Q}}_\ell) \times \mathbb{G}_m,$$

where $\varsigma : g \rtimes t \mapsto g\tilde{\rho}(t)$ is the isomorphism from Lemma 3.1.5.

Our goal is to verify (i) and (ii) of Theorem 3.2.7 for r^C. Since the codomain of r^C is identified with $\mathrm{GL}_N(\overline{\mathbb{Q}}_\ell) \rtimes (\{1, c\} \times \mathbb{G}_m)$ via $\tilde{\xi}_0$ above, we want to do the same with $\phi^C_{\pi_v} : W_{F_v} \to {}^C U_{F_v}$ via ${}^C U_{F_v} \cong \mathrm{GL}_N(\overline{\mathbb{Q}}_\ell) \rtimes (\{1, c\} \times \mathbb{G}_m)$ given by $\tilde{\xi}_u : {}^L U_{F_v} \cong \mathrm{GL}_N(\overline{\mathbb{Q}}_\ell) \rtimes \{1, c\}$ (and the identity map on the $\mathbb{G}_m$-factor of the C-group), which is consistent with $\tilde{\xi}_0$. For each $\sigma : F \hookrightarrow \mathbb{C}$, similarly $\zeta_{\pi,\sigma} \in X_*(\hat{T}_\sigma)_{\mathbb{Q}}$ is viewed as an element of $X_*(\mathbb{G}_m^N)_{\mathbb{Q}}$ via $\tilde{\xi}_{\tau_{0,\sigma}}$; see Case U of Section 2.1 for the discussions on $\tilde{\xi}_0$, $\tilde{\xi}_u$, and $\tilde{\xi}_{\tau_{0,\sigma}}$. Therefore (i) and (ii) are equivalent to the following assertions; see Section 2.1 for $\tau_{0,v}$ and $\tau_{0,\sigma}$:

(a) $\varsigma r^C|^{\mathrm{ss}}_{W_{F_v}} \cong \iota\tilde{\xi}_u(\phi^C_{\pi_v})$, for each finite place v of F not contained in S, and the place u of E induced by $\tau_{0,v} : E \hookrightarrow \overline{F}_v$.

(b) $\mu_{\varsigma r^C,\iota\sigma} = (-\iota\tilde{\xi}_{\tau_{0,\sigma}}(\zeta^C_{\pi,\sigma}), 1)$ for every embedding $\sigma : F \hookrightarrow \mathbb{C}$.

We observed that $\varsigma r^C = (R, \omega_\ell^{-1})$. Hence (a) holds after restriction to W_{E_u} by Proposition 3.2.4 (i”). Assertion (a) follows from this because the isomorphism class on each side is determined by its restriction to W_{E_u}; this is a special case of [Gan et al. 2012, Theorem 8.1(ii)]. As for (b), let $\tau_{0,\sigma} : E \hookrightarrow \mathbb{C}$ be as in Section 2.1, which extends σ. The Hodge–Tate cocharacters can be computed after taking a finite base extension, so

$$\mu_{\varsigma r^C,\iota\sigma} = \mu_{\varsigma r^C|_{\Gamma_E},\iota\tau_{0,\sigma}} = \mu_{(R,\omega_\ell^{-1}),\iota\tau_{0,\sigma}}.$$

Hence (b) is a consequence of Proposition 3.2.4(ii”) as well as the fact that ω_ℓ has Hodge–Tate weight -1. □

Acknowledgments. The idea to write this paper was conceived during a conversation with Tasho Kaletha, Peter Scholze, and others at the IHES summer school in July 2022. I am grateful to the organizers (Pierre-Henri Chaudouard, Wee Teck Gan, Tasho Kaletha, and Yiannis Sakellaridis) for creating the opportunity. I thank Alexander Bertoloni Meli, Toby Gee, and Alex Youcis for their comments on an earlier draft. Finally I am thankful to the referees for their valuable comments and suggestions.

References

[A’Campo 2024] L. A’Campo, “Rigidity of automorphic Galois representations over CM fields”, *Int. Math. Res. Not.* **2024**:6 (2024), 4541–4623. MR Zbl

[Arthur 2005] J. Arthur, “An introduction to the trace formula”, pp. 1–263 in *Harmonic analysis, the trace formula, and Shimura varieties*, edited by J. Arthur et al., Clay Math. Proc. **4**, Amer. Math. Soc., Providence, RI, 2005. MR Zbl

[Arthur 2013] J. Arthur, *The endoscopic classification of representations*, American Mathematical Society Colloquium Publications **61**, American Mathematical Society, Providence, RI, 2013. MR Zbl

[Arthur 2021] J. Arthur, "An introduction to Langlands functoriality", pp. 109–129 in *The genesis of the Langlands Program*, edited by J. Mueller and F. Shahidi, London Math. Soc. Lecture Note Ser. **467**, Cambridge Univ. Press, 2021. MR Zbl

[Arthur and Clozel 1989] J. Arthur and L. Clozel, *Simple algebras, base change, and the advanced theory of the trace formula*, Annals of Mathematics Studies **120**, Princeton University Press, 1989. MR Zbl

[Barnet-Lamb et al. 2014] T. Barnet-Lamb, T. Gee, D. Geraghty, and R. Taylor, "Potential automorphy and change of weight", *Ann. of Math.* (2) **179**:2 (2014), 501–609. MR Zbl

[Bellaïche and Chenevier 2009] J. Bellaïche and G. Chenevier, "Families of Galois representations and Selmer groups", *Astérisque* 324 (2009), xii+314. MR Zbl

[Bellaïche and Chenevier 2011] J. Bellaïche and G. Chenevier, "The sign of Galois representations attached to automorphic forms for unitary groups", *Compos. Math.* **147**:5 (2011), 1337–1352. MR Zbl

[Borel 1979] A. Borel, "Automorphic L-functions", pp. 27–61 in *Automorphic forms, representations and L-functions, II* (Proc. Sympos. Pure Math., Oregon State Univ., Corvallis, Ore., 1977), edited by A. Borel and W. Casselman, Proc. Sympos. Pure Math. **XXXIII**, Amer. Math. Soc., Providence, RI, 1979. MR Zbl

[Boxer and Pilloni 2021] G. Boxer and V. Pilloni, "Higher Coleman theory", preprint, 2021. Zbl arXiv 2110.10251

[Buzzard and Gee 2014] K. Buzzard and T. Gee, "The conjectural connections between automorphic representations and Galois representations", pp. 135–187 in *Automorphic forms and Galois representations, I* (Durham, 2011), edited by F. Diamond et al., London Math. Soc. Lecture Note Ser. **414**, Cambridge Univ. Press, 2014. MR Zbl

[Cai et al. 2019] Y. Cai, S. Friedberg, D. Ginzburg, and E. Kaplan, "Doubling constructions and tensor product L-functions: the linear case", *Invent. Math.* **217**:3 (2019), 985–1068. MR Zbl

[Caraiani and Le Hung 2016] A. Caraiani and B. V. Le Hung, "On the image of complex conjugation in certain Galois representations", *Compos. Math.* **152**:7 (2016), 1476–1488. MR Zbl

[Chaudouard and Laumon 2010] P.-H. Chaudouard and G. Laumon, "Le lemme fondamental pondéré, I: Constructions géométriques", *Compos. Math.* **146**:6 (2010), 1416–1506. MR Zbl

[Chaudouard and Laumon 2012] P.-H. Chaudouard and G. Laumon, "Le lemme fondamental pondéré, II: Énoncés cohomologiques", *Ann. of Math.* (2) **176**:3 (2012), 1647–1781. MR Zbl

[Clozel 1990] L. Clozel, "Motifs et formes automorphes: applications du principe de fonctorialité", pp. 77–159 in *Automorphic forms, Shimura varieties, and L-functions, I* (Ann Arbor, MI, 1988), edited by L. Clozel and J. S. Milne, Perspect. Math. **10**, Academic Press, Boston, 1990. MR Zbl

[Clozel et al. 2008] L. Clozel, M. Harris, and R. Taylor, "Automorphy for some l-adic lifts of automorphic mod l Galois representations", *Publ. Math. Inst. Hautes Études Sci.* 108 (2008), 1–181. MR Zbl

[Cogdell 2003] J. W. Cogdell, "Dual groups and Langlands functoriality", pp. 251–268 in *An introduction to the Langlands program* (Jerusalem, 2001), edited by J. Bernstein and S. Gelbart, Birkhäuser, Boston, 2003. MR Zbl

[Cogdell et al. 2011] J. W. Cogdell, I. I. Piatetski-Shapiro, and F. Shahidi, "Functoriality for the quasisplit classical groups", pp. 117–140 in *On certain L-functions*, edited by J. Arthur et al., Clay Math. Proc. **13**, Amer. Math. Soc., Providence, RI, 2011. MR Zbl

[Fakhruddin and Pilloni 2019] N. Fakhruddin and V. Pilloni, "Hecke operators and the coherent cohomology of Shimura varieties", preprint, 2019. To appear in *J. Inst. Math. Jussieu*. Zbl arXiv 1910.03790

[Gan et al. 2012] W. T. Gan, B. H. Gross, and D. Prasad, "Symplectic local root numbers, central critical L values, and restriction problems in the representation theory of classical groups", pp. 1–109 in *Sur les conjectures de Gross et Prasad*, *I*, Astérisque **346**, 2012. MR Zbl arXiv 0909.2999

[Harris et al. 2016] M. Harris, K.-W. Lan, R. Taylor, and J. Thorne, "On the rigid cohomology of certain Shimura varieties", *Res. Math. Sci.* **3** (2016), Paper No. 37, 308. MR Zbl

[Hevesi 2023] B. Hevesi, "Ordinary parts and local-global compatibility at $\ell = p$", 2023. arXiv 2311.13514

[Ishimoto 2023] H. Ishimoto, "The endoscopic classification of representations of non-quasi-split odd special orthogonal groups", 2023. Zbl arXiv 2301.12143

[Johansson 2013] C. Johansson, "A remark on a conjecture of Buzzard–Gee and the cohomology of Shimura varieties", *Math. Res. Lett.* **20**:2 (2013), 279–288. MR Zbl

[Johansson and Thorne 2020] C. Johansson and J. A. Thorne, "On subquotients of the étale cohomology of Shimura varieties", pp. 306–333 in *Shimura varieties*, edited by T. Haines and M. Harris, London Math. Soc. Lecture Note Ser. **457**, Cambridge Univ. Press, 2020. MR Zbl

[Kaletha et al. 2014] T. Kaletha, A. Minguez, S. W. Shin, and P.-J. White, "Endoscopic classification of representations: inner forms of unitary groups", 2014. Zbl arXiv 1409.3731

[Kottwitz and Shelstad 1999] R. E. Kottwitz and D. Shelstad, "Foundations of twisted endoscopy", *Astérisque* 255 (1999), vi+190. MR Zbl

[Kret and Shin 2020] A. Kret and S. W. Shin, "Galois representations for even general special orthogonal groups", 2020. To appear in *J. Inst. Math. Jussieu*. Zbl arXiv 2010.08408

[Kret and Shin 2023] A. Kret and S. W. Shin, "Galois representations for general symplectic groups", *J. Eur. Math. Soc.* (*JEMS*) **25**:1 (2023), 75–152. MR Zbl

[Labesse 2011] J.-P. Labesse, "Changement de base CM et séries discrètes", pp. 429–470 in *On the stabilization of the trace formula*, edited by L. Clozel et al., Stab. Trace Formula Shimura Var. Arith. Appl. **1**, International Press, Somerville, MA, 2011. MR Zbl

[Labesse and Lemaire 2021] J.-P. Labesse and B. Lemaire, "La formule des traces tordue pour les corps de fonctions", 2021. Zbl arXiv 2102.02517

[Lafforgue 2018] V. Lafforgue, "Chtoucas pour les groupes réductifs et paramétrisation de Langlands globale", *J. Amer. Math. Soc.* **31**:3 (2018), 719–891. MR Zbl

[Langlands 1970] R. P. Langlands, "Problems in the theory of automorphic forms", pp. 18–61 in *Lectures in Modern Analysis and Applications*, *III*, edited by C. T. Taam, Lecture Notes in Math. **Vol. 170**, Springer, 1970. MR Zbl

[Langlands and Shelstad 1987] R. P. Langlands and D. Shelstad, "On the definition of transfer factors", *Math. Ann.* **278**:1-4 (1987), 219–271. MR Zbl

[Lomelí 2009] L. A. Lomelí, "Functoriality for the classical groups over function fields", *Int. Math. Res. Not.* **2009**:22 (2009), 4271–4335. MR Zbl

[Mezo 2013] P. Mezo, "Character identities in the twisted endoscopy of real reductive groups", pp. vi+94 in Mem. Amer. Math. Soc. **1042**, Mem. Amer. Math. Soc., Providence, RI, 2013. MR Zbl

[Mœglin and Waldspurger 1989] C. Mœglin and J.-L. Waldspurger, "Le spectre résiduel de $\mathrm{GL}(n)$", *Ann. Sci. École Norm. Sup.* (4) **22**:4 (1989), 605–674. MR Zbl

[Mœglin and Waldspurger 2016a] C. Mœglin and J.-L. Waldspurger, *Stabilisation de la formule des traces tordue, I*, Progress in Mathematics **316**, Springer, 2016. MR Zbl

[Mœglin and Waldspurger 2016b] C. Mœglin and J.-L. Waldspurger, *Stabilisation de la formule des traces tordue, II*, Progress in Mathematics **317**, Springer, 2016. MR Zbl

[Mok 2015] C. P. Mok, *Endoscopic classification of representations of quasi-split unitary groups*, Mem. Amer. Math. Soc. **1108**, 2015. MR Zbl

[Patrikis 2015] S. Patrikis, "On the sign of regular algebraic polarizable automorphic representations", *Math. Ann.* **362**:1-2 (2015), 147–171. MR Zbl

[Rogawski 1990] J. D. Rogawski, *Automorphic representations of unitary groups in three variables*, Annals of Mathematics Studies **123**, Princeton University Press, 1990. MR Zbl

[Scholze 2015] P. Scholze, "On torsion in the cohomology of locally symmetric varieties", *Ann. of Math.* (2) **182**:3 (2015), 945–1066. MR Zbl

[Taïbi 2016] O. Taïbi, "Eigenvarieties for classical groups and complex conjugations in Galois representations", *Math. Res. Lett.* **23**:4 (2016), 1167–1220. MR Zbl

[Taïbi 2019] O. Taïbi, "Arthur's multiplicity formula for certain inner forms of special orthogonal and symplectic groups", *J. Eur. Math. Soc.* (*JEMS*) **21**:3 (2019), 839–871. MR Zbl

[Taïbi 2022] O. Taïbi, "Arthur parameters for automorphic representations of orthogonal groups", *Bull. Soc. Math. France* **150**:3 (2022), 473–516. Appendix to I. Chenevier, "Statistics for Kneser p-neighbors", Bull. Soc. Math. France **150**:3 (2022), 473–516. MR Zbl

[Taylor 2012] R. Taylor, "The image of complex conjugation in l-adic representations associated to automorphic forms", *Algebra Number Theory* **6**:3 (2012), 405–435. MR Zbl

[Waldspurger 2009] J.-L. Waldspurger, "À propos du lemme fondamental pondéré tordu", *Math. Ann.* **343**:1 (2009), 103–174. MR Zbl

[Waldspurger 2010] J.-L. Waldspurger, "Les facteurs de transfert pour les groupes classiques: un formulaire", *Manuscripta Math.* **133**:1-2 (2010), 41–82. MR Zbl

[Zhu 2020a] X. Zhu, "Coherent sheaves on the stack of Langlands parameters", 2020. Zbl arXiv 2008.02998

[Zhu 2020b] X. Zhu, "A note on Integral Satake isomorphisms", 2020. Zbl arXiv 2005.13056

Received 13 Feb 2023. Revised 28 Mar 2024.

SUG WOO SHIN:

sug.woo.shin@berkeley.edu
Department of Mathematics, UC Berkeley, Berkeley, CA, United States

and

Korea Institute for Advanced Study, Seoul, Republic of Korea

msp

Essential Number Theory
Vol. 3, No. 1, 2024
https://doi.org/10.2140/ent.2024.3.63

L-values and nonsplit extensions: a simple case

Christopher Skinner

We explain a construction of explicit extensions — of rational Hodge structures and of p-adic Galois representations — in a simple context: the cohomology of $\mathbb{P}^1 - \{\text{some points}\}$ relative to {some other points}. These extensions are naturally related to Dirichlet characters, and we connect the nonsplitting of these extensions to the values at $s = 0$ and $s = 1$ of associated Dirichlet L-functions $L(s, \chi)$. We highlight the close parallels between the proofs of nonsplitting in both the Hodge-theoretic and p-adic cases, emphasizing the use of de Rham theory. We also indicate connections with Euler systems along with variations on these constructions in the setting of modular curves. This paper is intended as an introduction to some of the key ideas in forthcoming constructions of Galois cohomology classes and Euler systems in a range of settings.

1. Introduction

Beginning with Birch and Swinnerton-Dyer's formulation of their celebrated conjecture, if not earlier, number theorists have sought arithmetic explanations for the zeros at special values of s of the L-functions $L(M, s)$ that arise in the context of arithmetic geometry. This encompasses Dirichlet L-series, L-functions of algebraic Hecke characters, the Hasse–Weil L-functions of elliptic curves and other varieties over number fields, etc. For example, conjectures of Beilinson essentially express the order of vanishing at particular s as the dimension of a certain group of extensions

MSC2020: 11F80.
Keywords: L-values, Galois extensions, mixed Hodge structures, Euler systems.

in a category of mixed Hodge structures, or more ambitiously in a category of mixed motives [Beilinson 1984; Nekovář 1994]. And conjectures of Bloch and Kato essentially express the same orders of vanishing as the dimensions of certain groups of extensions of p-adic Galois representations [Bloch and Kato 1990; Fontaine and Perrin-Riou 1994]. The latter should be the p-adic realizations of the former motivic extensions. These conjectures are only proved for some simple cases, though evidence exists for many interesting L-functions. It is expected that the Galois extensions related to a given $L(M, s)$ and its twists $L(M, \chi, s)$ by Dirichlet characters (or other finite Hecke characters) should form an Euler system, which then yield — via the theory of Euler and Kolyvagin systems — upper bounds on orders of related Selmer groups.

Given an L-function $L(M, s)$ and a special value of s, the expected motivic nature of the related extensions makes it natural to ask: should the expected extensions be concretely realized in cohomology by some general construction? A good rule of thumb here is that if $L(M, s)$ is suitably primitive and indecomposable, then this should be the case if and only if the order of vanishing equals 1: there is generally no good reason to distinguish one line in a space of extensions from another when the space of extensions has dimension greater than 1. Such a guiding principle both explains and predicts the extensions that comprise many of the known examples of Euler systems.[1]

We construct explicit extensions — of rational Hodge structures and of p-adic Galois representations — in a simple context: the cohomology of $\mathbb{P}^1 - \{\text{some points}\}$ relative to $\{\text{some other points}\}$. These extensions are extensions in the corresponding categories, that is, elements of Ext^1-groups. They are naturally related to Dirichlet characters χ, and for nontrivial χ we demonstrate that they are nonsplit if and only if χ is even and $L(s, \chi)$ vanishes at $s = 0$ to order 1. Our aim in writing this is three-fold: (i) to provide some evidence in a very simple case for the rule of thumb stated above, (ii) to highlight the close parallels between the proofs of nonsplitting in both the Hodge-theoretic and p-adic cases, and (iii) to give a sense, in this very simple case, of the ideas underpinning some recent and forthcoming constructions of new Euler systems (such as [Sangiovanni-Vincentelli and Skinner $\geq$ 2024a; $\geq$ 2024b], but see also [Shang et al. $\geq$ 2024]). We emphasize especially the aim (ii), though we also provide some elaboration on (iii).

In both the Hodge and p-adic cases, the proof of nonsplitting is reduced to an analytic calculation. For the Hodge structures this goes via Hodge theory and the real analytic de Rham isomorphism. For the p-adic Galois representations this goes via the comparison isomorphisms of p-adic Hodge theory as well as a p-adic analytic expression for algebraic de Rham classes. In our simple setting we can appeal to

[1] But like all such ‘rules’, it should also be taken with a grain of salt.

Monsky–Washnitzer cohomology for the latter, though the final calculation is done in the context of locally analytic functions via Coleman's p-adic integration. In both cases, the crucial input is a simple, explicit description of a cohomology class and its de Rham realization. In fact, another of the key points to carry away from this note is that — at least for the purposes of Euler systems — in many instances such explicit classes can reasonably substitute for motivic constructions of classes (often realized via, say, units or elements of higher Chow groups).

The constructions in our simple case are carried out in Sections 4 and 5. The aim in each section is an explanation of the statements (4.6.b) and (5.6.c), respectively, linking orders of zeros of complex L-functions to the nontriviality of extensions. We also indicate the connection with Euler systems in Sections 5.7 and 5.8, respectively. This is followed in Section 6 with brief sketches of similar constructions and calculations in the cohomology of modular curves, yielding extensions related to L-values of Dirichlet characters (again) and to Hecke characters of imaginary quadratic fields.

We suspect that many of the ideas herein, especially in the simple context in which we work, are known to experts.[2] However, extracting them from the more well-known of the existing literature (such as [Deligne 1989] or [Deligne and Goncharov 2005]) does not seem straightforward, which hopefully lends some usefulness to publishing this note. At the end of Section 3 we give some indication of the relation of this note to other works. Of course, our goal is not so much to prove new results but to explain old results from a perspective that might not be widely known.

2. The setting

Let $X = \mathbb{P}^1_{/\mathbb{Q}} = \operatorname{Proj} \mathbb{Q}[t_0, t_1]$. Let $\infty \in X(\mathbb{Q})$ be the point $\infty = [0 : 1]$. Let $\mathbb{A}^1 = \mathbb{P}^1 \setminus \{\infty\}$, so $\mathbb{A}^1 = \operatorname{Spec} \mathbb{Q}[t]$, $t = t_1/t_0$. Let $Y = \mathbb{A}^1 \setminus \{1\} = \operatorname{Spec} \mathbb{Q}\big[t, \frac{1}{t-1}\big]$. So $Y = X \setminus Z$ for $Z = \{\infty, 1\}$. Let $N \geq 2$ be an integer and let $W = \mu_N^\circ = \operatorname{Spec} \mathbb{Q}[t]/(\Phi_N(t)) \subset \mathbb{A}^1$, for $\Phi_N(t)$ the N-th cyclotomic polynomial. In particular, $X(\mathbb{C}) = \mathbb{P}^1(\mathbb{C})$ is just the Riemann sphere, $Z(\mathbb{C}) = \{\infty, 1\}$, $Y(\mathbb{C}) = X(\mathbb{C}) \setminus Z(\mathbb{C})$ is just the punctured plane $\mathbb{C} \setminus \{1\}$, and $W(\mathbb{C}) = \{\exp(2\pi i a/N) : a \in (\mathbb{Z}/N\mathbb{Z})^\times\}$ is the set of primitive N-th roots of unity. Since $N \geq 2$, $1 \notin W(\mathbb{C})$.

3. The basic idea

To construct and analyze the extensions in this paper we will make use of various cohomology theories for X, Y, Z, and W: the singular and de Rham cohomologies of the manifolds defined by the $\mathbb{C}$-points of these varieties, the étale and algebraic de Rham cohomologies of the varieties, and even crystalline cohomology. Each

[2] Harder's unpublished manuscript [2023], especially §2, provides clear evidence of this.

of these cohomology theories admits relative cohomology for the pairs (X, Y) and (Y, W), yielding exact sequences

$$\cdots \to H^i(X) \to H^i(Y) \to H^{i+1}(X, Y) \to H^{i+1}(X) \to \cdots$$

and

$$\cdots \to H^{i-1}(W) \to H^i(Y, W) \to H^i(Y) \to H^i(W) \to \cdots .$$

Here we have written $H^i(-)$ to denote any of the cohomology theories and have suppressed any reference to coefficients (which may depend on the particular theory). Purity often affords a canonical identification of $H^{i+1}(X, Y)$ (which is sometimes written as $H^{i+1}_Z(X)$) with $H^{i-1}(Z)(-1)$, where the (-1) denotes a twist, whose nature depends on the cohomology theory (e.g., a Tate twist in the case of étale cohomology). We also refer to the first of these sequences as the Gysin sequence for the pair (X, Z).

We will use the first of the above sequences together with purity to define explicit submodules[3] $A \subset H^1(Y)$ and to deduce various properties of A (e.g., the Galois action on A in the case of étale cohomology). We will also define an explicit quotient $H^0(W) \twoheadrightarrow B$ that factors through $H^0(W)/\mathrm{im}(H^0(Y))$. We will then use the second of the above sequences to define an extension via pull-back/push-forward:

$$\begin{tikzcd} \frac{H^0(W)}{\mathrm{im}(H^0(Y))} \arrow[r, hook] \arrow[d, two heads] & H^1(Y, W) \arrow[r, two heads] \arrow[d, dashed] & H^1(Y) \\ B \arrow[r, hook] & \mathcal{E} \arrow[r, two heads] & A. \arrow[u, hook] \end{tikzcd}$$

Here the dashed arrow denotes subquotient. The particular category in which the extension $\mathcal{E}$ belongs depends on the cohomology theory (e.g., the category of Galois modules in the case of étale cohomology). Our aim is to understand when the extension class $\mathcal{E} \in \mathrm{Ext}^1(A, B)$ is nonzero, that is, when the extension $\mathcal{E}$ is nonsplit. This will be achieved by making use of the comparison isomorphisms of the various cohomology theories, which will ultimately reduce the problem to whether a certain formula extracted from de Rham cohomology is nonzero.

A quick glance at a select part of the literature. We very briefly indicate the relation of the construction sketched above to some of the vast body of literature about mixed motives.

1) (Nori motives). Our use of relative cohomology meshes well with Nori's program to construct a general theory of mixed motives using such cohomology groups. A nice exposition of Nori's program is given in [Huber and Müller-Stach 2017]. Not surprisingly there is some overlap of the context we work in with some of the

[3]In the simplest situation considered here, A will be turn out to be all of $H^1(Y)$, but more generally it will just be a submodule (see 5.8).

examples in op. cit., especially [Huber and Müller-Stach 2017, §14.1]. However, the emphasis therein, as in the complementary survey [Huber 2020], is on periods, while the focus herein is on showing that certain explicit extensions of motives, and especially of Galois representations, are nonsplit. Of course, the calculations in Sections 4.6 and 5.6 can be recast in the context of periods and the final results expressed as: certain periods are nonzero if and only if certain extensions are nonsplit (the motivated reader might profit from doing so).

2) ($\mathbb{P}^1$ minus three points) Deligne's influential paper [1989] introduced, among other things, an approach to studying the category of mixed Tate motives (the kind of extensions we construct in this note) via the unipotent fundamental group of $X = \mathbb{P}^1 - \{0, 1, \infty\}$; this was realized more completely in [Deligne and Goncharov 2005]. This essentially realizes extensions of Tate motives in the cohomology of X^n relative to certain normal crossing divisors (which, in the cases considered, can be reinterpreted as being in the cohomology of certain unipotent local systems on X); see especially [Deligne and Goncharov 2005, §3]. This setting is well-adapted for expressing associated periods as iterated integrals (hence the relation to multiple zeta values; see [Brown 2014] for the state of the art). A translation of the construction we explain herein into the setting of [Deligne 1989; Deligne and Goncharov 2005] would surely be interesting, but we content ourselves with noting that the *duals* of the extensions we construct in Sections 4 and 5 can be extracted from the special case of $X = \mathbb{P}^1 - \mu_N^\circ$, $\{a, b\} = \{1, \infty\}$, and $n = 1$ (see [Deligne and Goncharov 2005, Proposition 3.4]).

3) (Harder's Anderson motives). After preparing the first draft of this note we became aware of an unpublished manuscript of Harder [2023] in which he proposes a very similar construction of mixed motives and Hodge structures, which he calls Anderson motives. Indeed, our construction can be viewed as an elucidation of a special case of Harder's construction for curves [2023, §2]. One thing this note includes that is not in op. cit. is an explanation of the nonsplitting of the p-adic Galois representations. Indeed, illustrating how the arguments for Galois representations closely parallel those for the Hodge structures is one of our main points. We also explain — at least in our simple case, but see also [Shang et al. $\geq$ 2024] or [Sangiovanni-Vincentelli and Skinner $\geq$ 2024a] — how these constructions lead to Euler systems, which answers questions raised by Harder.

4) (Beilinson's conjectures) The extensions that we construct — of mixed Hodge structures and of p-adic Galois representations — are shown to be nontrivial precisely when the value of some Dirichlet L-series $L(s, \bar{\chi})$ is nonzero at $s = 1$. By the functional equation, this can be reinterpreted as saying that $\mathrm{ord}_{s=0} L(s, \chi) = 1$. Very generally, Beilinson conjectured that the order of vanishing of an L-function $L(s, M)$ of a motive M at the special value $s = 0$ should (usually) equal the

rank of the group of extensions $\mathrm{Ext}^1_{MM}(\mathbb{Q}(0), M^\vee(1))$, in the category MM of mixed motives, of the trivial Tate motive $\mathbb{Q}(0)$ by the dual motive $M^\vee(1)$. He also conjecture an expression for a certain associated regulator map in terms of the first nonzero coefficient of the Taylor series of $L(s, M)$ at $s = 0$. The expository article [Nekovář 1994] is an excellent introduction to these conjectures. In this paper we essentially construct extensions $\mathbb{Q}_{\bar{\chi}} \hookrightarrow E \twoheadrightarrow \mathbb{Q}(-1)$ for a motive $\mathbb{Q}_{\bar{\chi}}$ associated with $\bar{\chi}$ (what we really construct should be the Hodge and p-adic étale realizations of such motivic extensions). Then $E(1) \in \mathrm{Ext}^1_{MM}(\mathbb{Q}(0), \mathbb{Q}_{\bar{\chi}}(1))$. Beilinson's conjectures tell as that the right-hand side should be nonzero if and only if $L(0, \chi) = 0$ (as $\mathbb{Q}_{\bar{\chi}}(1)$ is the dual of $\mathbb{Q}_\chi$ and $L(s, \mathbb{Q}_\chi) = L(s, \chi)$), and we show that if χ is even and $\mathrm{ord}_{s=0} L(s, \chi) = 1$ then $E(1) \neq 0$. Of course, Dirichlet's units theorem already tells us that the rank of $\mathrm{Ext}^1_{MM}(\mathbb{Q}(0), \mathbb{Q}_\chi(1))$ is 1 if χ is even (see [Nekovář 1994, §8, (2)]). Our focus is on showing that a *particular* construction yields a nonsplit extension.

4. Nonsplit extensions of rational Hodge structures

We find nonsplit extensions of rational Hodge structures in the relative cohomology of the pair (Y, W). We check that these extensions are nonsplit essentially by integrating an explicit differential representing a class in the cohomology of Y and recognizing the resulting formulas as expressions for L-values of Dirichlet characters at $s = 1$ (or derivatives at $s = 0$ via the functional equation). The key input here is the explicit de Rham representative of the cohomology class.

Though the idea is simple — and the integration boils down to $\frac{dx}{x} = d \log |x|!$ — we have included details of the singular and de Rham cohomology of Y and the pair (Y, W). We have done this partly for the sake of completeness, partly to illustrate the general definitions in this simple case, and partly to more clearly set out a template for other situations (see Section 6). A reader with some familiarity with mixed Hodge structures should be able to grasp the gist quickly upon reading Section 4.2 and fill in details by scanning the subsequent displayed equations. For readers less familiar with Hodge theory, we have included a brief discussion and description of the main players and tried to point to some useful resources, particularly in Sections 4.1, 4.3, and 4.4.

Conventions. In the following, given a variety $\mathcal{V}$ over a subfield of $\mathbb{C}$ and a field F we write $H^i(\mathcal{V}, F)$ for the singular cohomology group $H^i(\mathcal{V}(\mathbb{C}), F)$, and similarly for relative cohomology with respect to a subvariety $\mathcal{U}$ of $\mathcal{V}$. For nonsingular $\mathcal{V}$ and $\mathcal{U}$ these cohomology groups are canonically computed by de Rham cohomology (which gives rise to the Hodge filtrations on the former), and the latter is computed by the real-analytic de Rham complex and by the hypercohomology of both the holomorphic and algebraic de Rham complexes (we abuse notation by not distinguishing our notation for the latter two). We write ι_{dR} for these de Rham-singular isomorphisms. They are functorial in $\mathcal{V}$ and compatible with the long-exact sequences for relative cohomology, Gysin sequences, etc.

Let $\mathbb{Z}(1) = 2\pi i\mathbb{Z}$ and $\mathbb{Z}(n) = \mathbb{Z}(1)^{\otimes n}$ for any integer n. Note that $\mathbb{Z}(-1)$ is canonically identified with $(2\pi i)^{-1}\mathbb{Z}$. We write $H^i(V, F)(n)$ to mean $H^i(V, F) \otimes \mathbb{Z}(n)$. Keeping track of such 'twists' makes comparisons with de Rham and étale cohomology more clearly functorial. The Hodge filtration on $H^i(V, \mathbb{C})(n)$ comes from that of $H^i(V, \mathbb{C})$ with the index shifted by $+n$, and the weight filtration is also the same but with index shifted by $+2n$, and likewise for relative cohomology. Similarly, our conventions for twists of de Rham cohomology are such that $H^i_{\mathrm{dR}}(V/F)(n)$ is $H^i_{\mathrm{dR}}(V/F)$ with the Hodge filtration shifted by $+n$ and the weight filtration by $+2n$.

4.1. ***Hodge structures and extensions, briefly.*** Recall that a rational mixed Hodge structure is a finite-dimensional $\mathbb{Q}$-space V together with:

- (Hodge filtration) a decreasing filtration $\cdots \supseteq F^p V_{\mathbb{C}} \supseteq F^{p+1} V_{\mathbb{C}} \supseteq \cdots$ of the complex vector space $V_{\mathbb{C}} = V \otimes \mathbb{C}$ such that $F^p V_{\mathbb{C}} = V_{\mathbb{C}}$ if $p \ll 0$ and $F^p V_{\mathbb{C}} = 0$ if $p \gg 0$, and
- (weight filtration) an increasing filtration $\cdots \subseteq W_n V \subseteq W_{n+1} V \subseteq \cdots$ of the rational vector space V such that $W_n V = V$ if $n \gg 0$ and $W_n V = 0$ if $n \ll 0$,

that satisfy:

- (pure graded pieces) the filtration $F^p V_{\mathbb{C}}$ induces a filtration on $\mathrm{gr}_n V_{\mathbb{C}}$ for $\mathrm{gr}_n V = W_n V / W_{n-1} V$,

$$F^p(\mathrm{gr}_n V_{\mathbb{C}}) = (F^p V_{\mathbb{C}} \cap W_n V_{\mathbb{C}} + W_{n-1} V_{\mathbb{C}}) / W_{n-1} V_{\mathbb{C}},$$

and

$$\mathrm{gr}_n V_{\mathbb{C}}^{p,q} := F^p(\mathrm{gr}_n V_{\mathbb{C}}) \cap \overline{F^q(\mathrm{gr}_n V_{\mathbb{C}})}$$

is such that $\mathrm{gr}_n V_{\mathbb{C}}^{p,q} = 0$ if $p + q \neq n$ and

$$\mathrm{gr}_n V_{\mathbb{C}} = \bigoplus_{p+q=n} \mathrm{gr}_n V_{\mathbb{C}}^{p,q}.$$

Here the overline $\overline{(\cdot)}$ denotes the image under the action of complex conjugation on the scalars of $V_{\mathbb{C}} = V \otimes \mathbb{C}$.

A (mixed) Hodge structure with graded weight filtration supported on exactly one degree, that is, $\mathrm{gr}_n V = V$ for some (unique) n, is a *pure* Hodge structure of weight n. So the third condition above just says that for a mixed Hodge structure, the induced Hodge structures on the graded pieces of the weight filtration are pure of the corresponding weight. A morphism of mixed Hodge structures is a $\mathbb{Q}$-linear map that is compatible with the Hodge and weight filtrations. Let $\mathbb{Q}-\mathrm{MHS}$ denote the category of mixed rational Hodge structures. Replacing $\mathbb{Q}$ with $\mathbb{R}$ in the above, we get the category $\mathbb{R}-\mathrm{MHS}$ of real mixed Hodge structures. A rational mixed Hodge structure V gives rise to a real mixed Hodge structure $V_{\mathbb{R}} = V \otimes \mathbb{R}$ by extending scalars.

The singular cohomology groups of an algebraic variety (including relative cohomology) are all equipped with canonical rational mixed Hodge structures, and all the maps in the associated long exact sequences (e.g., the Gysin sequence and the sequence for relative cohomology) are morphisms of mixed Hodge structures. This is a consequence of Hodge theory as developed in [Deligne 1971a; 1971b; 1974]; see also [Peters and Steenbrink 2008]. The article [Kedlaya 2008] contains a fairly gentle introduction to Hodge theory for varieties.

The simplest examples of nonzero Hodge structures are the pure Hodge structures $\mathbb{Q}(m) = (2\pi i\mathbb{Q})^{\otimes m} = (2\pi i)^m\mathbb{Q}$, m an integer, with Hodge filtration $F^p\mathbb{Q}(m) = \mathbb{Q}(m)$ if $p \le -m$ and $F^p\mathbb{Q}(m) = 0$ if $p > -m$ and weight filtration $W_n\mathbb{Q}(m) = \mathbb{Q}(m)$ if $n \ge -2m$ and $W_n\mathbb{Q}(m) = 0$ if $n < -2m$. So $\mathbb{Q}(m)$ is pure of weight $-2m$ and $\mathbb{C}(m) = \mathbb{Q}(m) \otimes \mathbb{C} = \mathbb{Q}(m)_{\mathbb{C}} = \mathbb{C}(m)^{-m,-m}$. If $\mathcal{V}$ is a complete, connected variety over $\mathbb{C}$ of dimension d, then $H^{2d}(\mathcal{V}, \mathbb{Q})$ is isomorphic to $\mathbb{Q}(-d)$ as a Hodge structure. Let $\mathbb{R}(m) = \mathbb{Q}(m) \otimes \mathbb{R}$; this is a pure real Hodge structure.

The simplest examples of nontrivial rational mixed Hodge structures are the nonsplit extensions

$$0 \to \mathbb{Q}(n) \to E \to \mathbb{Q}(m) \to 0$$

of $\mathbb{Q}(m)$ by $\mathbb{Q}(n)$, $m < n$, in the category of mixed Hodge structures.[4] Let $\phi_H : \mathbb{C}(m) \to E_{\mathbb{C}}$ be a $\mathbb{C}$-linear splitting compatible with the Hodge filtrations; ϕ_H is unique in this case. Let $\phi_W : \mathbb{Q}(m) \to E$ be a $\mathbb{Q}$-linear splitting respecting the weight filtrations; in this case, ϕ_W is only well-defined up to addition of any element of $\mathrm{Hom}_{\mathbb{Q}}(\mathbb{Q}(m), \mathbb{Q}(n))$. Let $\phi = \phi_H - \phi_W \in \mathrm{Hom}_{\mathbb{C}}(\mathbb{C}(m), \mathbb{C}(n))$. Then the image of ϕ in $\mathrm{Hom}_{\mathbb{C}}(\mathbb{C}(m), \mathbb{C}(n))/\mathrm{Hom}_{\mathbb{Q}}(\mathbb{Q}(m), \mathbb{Q}(n))$ depends only on E and not the choices of ϕ_H or ϕ_W. This yields an identification $\mathrm{Ext}^1_{\mathbb{Q}-\mathrm{MHS}}(\mathbb{Q}(m), \mathbb{Q}(n)) = \mathrm{Hom}_{\mathbb{C}}(\mathbb{C}(m), \mathbb{C}(n))/\mathrm{Hom}_{\mathbb{Q}}(\mathbb{Q}(m), \mathbb{Q}(n)) \simeq \mathbb{C}/\mathbb{Q}$. Injectivity is a consequence of the observation that E is split if and only if we can choose $\phi_W = \phi_H$, so E is split if and only if $\phi_H - \phi_W \in \mathrm{Hom}_{\mathbb{Q}}(\mathbb{Q}(m), \mathbb{Q}(n))$ in general. Surjectivity follows by an explicit construction: Let $\phi : \mathbb{C}(m) \to \mathbb{C}(n)$ be a $\mathbb{C}$-linear map (which necessarily preserves the weight filtrations in this case). Consider the vector space $E = \mathbb{Q}(n) \oplus \mathbb{Q}(m)$ with Hodge and weight filtrations

$$F^pE_{\mathbb{C}} = \{(a + \phi(b), b) : a \in F^p\mathbb{C}(n), b \in F^p\mathbb{C}(m)\},$$
$$W_kE = \begin{cases} E, & k \ge -m, \\ \mathbb{Q}(n), & -n \le k < -m, \\ 0, & k < -n. \end{cases}$$

[4] If $n \le m$ then any such extension is split.

This is a mixed Hodge structure, and the natural inclusion $\mathbb{Q}(n) \hookrightarrow E$ and projection $E \twoheadrightarrow \mathbb{Q}(m)$ are clearly morphisms of Hodge structures: $E \in \mathrm{Ext}^1_{\mathbb{Q}-\mathrm{MHS}}(\mathbb{Q}(m), \mathbb{Q}(n))$. The image of this extension in $\mathrm{Hom}_{\mathbb{C}}(\mathbb{C}(m), \mathbb{C}(n))/\mathrm{Hom}_{\mathbb{Q}}(\mathbb{Q}(m), \mathbb{Q}(n))$ is just the image of ϕ. Similarly, $\mathrm{Ext}^1_{\mathbb{R}-\mathrm{MHS}}(\mathbb{R}(n), \mathbb{R}(n))$ is identified with the space $\mathrm{Hom}_{\mathbb{C}}(\mathbb{C}(m), \mathbb{C}(n))/\mathrm{Hom}_{\mathbb{R}}(\mathbb{R}(n), \mathbb{R}(m)) \simeq \mathbb{C}/\mathbb{R}$ and so is a one-dimensional $\mathbb{R}$-space. For more on extensions of mixed Hodge structure, the interested reader should consult [Carlson 1980] or [Carlson and Hain 1989].

In the rest of Section 4 we will find extensions of the Hodge structures $\mathbb{Q}(-1)$ by $\mathbb{Q}(0)$ as quotients of the relative singular cohomology groups $H^1(Y, W, \mathbb{Q})$ of the pairs (Y, W), and so find extensions in $\mathbb{Q}-\mathrm{MHS}$. To decide whether such an extension E is nontrivial, it suffices to identify a homomorphism $\phi : \mathbb{C}(-1) \to \mathbb{C}(0)$ giving rise to E as above. One way of doing this is as follows: Let $0 \neq \omega \in \mathbb{Q}(-1)$ and find elements $\omega_H \in F^1 E_{\mathbb{C}}$ and $\omega_W \in W_2 E$ that both map to ω. Identifying $\mathbb{Q}(-1)$ with a subspace of E via $\omega \mapsto \omega_W$, the Hodge structure on E is identified with the Hodge structure on $\mathbb{Q}(0) \oplus \mathbb{Q}(-1)$ defined by the ϕ such that $\phi(\omega) = \omega_H - \omega_W$. This extension is split if and only if $\phi \in \mathrm{Hom}_{\mathbb{Q}}(\mathbb{Q}(m), \mathbb{Q}(n)))$ and so if and only if $\omega_H - \omega_W \in \mathbb{Q}(0)$. If we work instead in $\mathbb{R}-\mathrm{MHS}$, then the criteria to be split becomes $\omega_H - \omega_W \in \mathbb{R}(0)$. In practice (as will be the case below) one can often find an explicit ω_H using Hodge theory, but finding an explicit ω_W — especially one for which the difference $\omega_H - \omega_W$ can be identified — can be difficult. We take a different tack.

To prove nontriviality of the extensions we find, we make use of the fact that all the Hodge structures involved in our constructions have a particular enrichment. This enrichment is the action of an involution ϕ_∞ on the underlying $\mathbb{Q}$-space V of a rational mixed Hodge structure (or $\mathbb{R}$-space of a real mixed Hodge structure) such that the action of $\phi_\infty \otimes \tau$ on $V_{\mathbb{C}} = V \otimes \mathbb{C}$ induces a $\mathbb{C}$-semilinear involution of each $F^p V_{\mathbb{C}}$. Here τ denote the action of complex conjugation on $\mathbb{C}$ and the semilinearity is with respect to τ. We denote by $\mathbb{Q}-\mathrm{MHS}^+$ the category of such enriched rational mixed Hodge structures (morphisms must also respect the action of ϕ_∞). We similarly write $\mathbb{R}-\mathrm{MHS}^+$ for the category of such enriched real mixed Hodge structures. The Hodge structures coming from the singular cohomology of varieties defined over $\mathbb{R}$ or some subfield have a natural enrichment: ϕ_∞ is the involution induced from the action of complex conjugation on the $\mathbb{C}$-points of the variety. The Hodge structures $\mathbb{Q}(m)$ also have natural enrichments: ϕ_∞ acts as multiplication by $(-1)^m$. For a complete, geometrically connected variety $\mathcal{V}$ of dimension d defined over a subfield of $\mathbb{R}$, the enriched Hodge structure on $H^{2d}(\mathcal{V}, \mathbb{Q})$ is isomorphic to that of $\mathbb{Q}(-d)$.

Let $m < n$ be integers. Following the description of $\mathrm{Ext}^1_{\mathbb{Q}-\mathrm{MHS}}(\mathbb{Q}(m), \mathbb{Q}(n))$ above, we find that extensions in $\mathrm{Ext}^1_{\mathbb{Q}-\mathrm{MHS}^+}(\mathbb{Q}(m), \mathbb{Q}(n))$ are those coming from homomorphisms $\phi \in \mathrm{Hom}_{\mathbb{C}}(\mathbb{C}(m), \mathbb{C}(n))$ such that $\phi((2\pi i)^m) = r i^{m-n}(2\pi i)^n$ for some $r \in \mathbb{R}$. In particular, the group of enriched extensions $\mathrm{Ext}^1_{\mathbb{Q}-\mathrm{MHS}^+}(\mathbb{Q}(m), \mathbb{Q}(n))$ is identified with the image of $i\mathbb{R}$ in $\mathbb{C}/\mathbb{Q}$ if $n-m$ is odd and with $\mathbb{R}/\mathbb{Q}$ otherwise. Similarly, $\mathrm{Ext}^1_{\mathbb{R}-\mathrm{MHS}^+}(\mathbb{R}(m), \mathbb{R}(n))$ is identified with $i\mathbb{R} \xrightarrow{\sim} \mathbb{C}/\mathbb{R}$ if $m-n$ is odd, but $\mathrm{Ext}^1_{\mathbb{R}-\mathrm{MHS}^+}(\mathbb{R}(m), \mathbb{R}(n)) = 0$ if $m-n$ is even.

Our strategy for determining whether an extension E of $\mathbb{Q}(-1)$ by $\mathbb{Q}(0)$ in $\mathbb{Q}-\mathrm{MHS}^+$ is nonzero will be to find an explicit element $0 \neq \omega \in \mathbb{Q}(-1)$ and an explicit lift $\omega_H \in F^1E_{\mathbb{C}}$. Then E is nonsplit if and only if $\phi_\infty(\omega_H)+\omega_H \neq 0$. This is readily seen by using the description of E as an extension associated with some $\phi \in \mathrm{Hom}_{\mathbb{C}}(\mathbb{C}(-1), \mathbb{C}(0))$ such that $\phi((2\pi i)^{-1}) = ir \in i\mathbb{R}$ (which is split if and only if $r=0$). For if E is isomorphic to the enriched mixed Hodge structure on $\mathbb{Q}(0) \oplus \mathbb{Q}(-1)$ for some ϕ such that $\phi((2\pi i)^{-1}) = ir \in i\mathbb{R}$ and if $\omega = (2\pi i)^{-1}a$, then $\omega'_H = (ira, (2\pi i)^{-1}a) \in F^1E_{\mathbb{C}}$ is a lift of ω and $\phi_\infty(\omega'_H)+\omega'_H = 2ira \in \mathbb{C}(0)$ is nonzero if and only if $r \neq 0$, that is, if and only if $\phi \neq 0$, which – as we have already seen – is the condition for E to be nonsplit in $\mathbb{Q}-\mathrm{MHS}^+$. As ϕ_∞ acts trivially, on $\mathbb{Q}(0)$ and hence on $\mathbb{C}(0)$, we see that $\phi_\infty(\omega_H)+\omega_H = \phi_\infty(\omega'_H)+\omega'_H$. Of course, the 'if' part is even easier to see in this special case: ω_H is the unique lift of ω to $F^1E_{\mathbb{C}}$ and so the extension being split in $\mathbb{Q}-\mathrm{MHS}^+$ would then imply that $\phi_\infty\omega_H = -\omega_H$. In practice, we will be able to use Hodge theory to explicitly compute $\phi_\infty(\omega_H)+\omega_H$. In this approach we only make use of an explicit lift $\omega_H \in F^1E_{\mathbb{C}}$ and do not need to also identify a lift $\omega_W \in W_2E$. As is explained in Section 5, a very similar strategy can be used to show that certain extensions of p-adic Galois representations are nonsplit (see especially Section 5.1).

4.2. *The extension $\mathcal{E}_{\mathrm{MH}}$.* Let $F/\mathbb{Q}$ be any extension. The relative singular cohomology $H^i(Y, W, F)$ fits into a long exact sequence

$$\cdots \to H^0(Y,F) \to H^0(W,F) \to H^1(Y,W,F) \to H^1(Y,F) \to H^1(W,F) \to \cdots. \tag{4.2.a}$$

In the case $F = \mathbb{Q}$, each of the cohomology groups in this sequence is endowed with a rational (possibly mixed) Hodge structure, and the maps between groups are morphisms of mixed Hodge structures. In this case, the Hodge structures on $H^0(W, F)$ and $H^1(Y, F)$ are pure of weights 0 and 2, respectively (for more details on the cohomology and Hodge theory of Y and (Y, W) and associated notation, see Sections 4.3 and 4.4 below). In particular, the induced extension

$$0 \to \frac{H^0(W,\mathbb{Q})}{\mathrm{im}(H^0(Y,\mathbb{Q}))} \to H^1(Y,W,\mathbb{Q}) \to H^1(Y,\mathbb{Q}) \to 0 \tag{4.2.b}$$

realizes the mixed Hodge structure on $H^1(Y, W, \mathbb{Q})$ as an extension in the category $\mathbb{Q}-\mathrm{MHS}$ of mixed rational Hodge structures: an extension of a pure Hodge structure of weight 0 by a pure Hodge structure of weight 2. Since each of the varieties X,

Y, Z, and W is defined over $\mathbb{Q}$, the singular cohomology groups considered above all carry the action of an involution, denoted ϕ_∞, induced by the action of complex conjugation on the $\mathbb{C}$-points of the varieties, and the above maps of cohomology groups also respect the action of ϕ_∞. The tensor product $\phi_\infty \otimes \tau$ of ϕ_∞ with the action τ of complex conjugation on the coefficients for $F = \mathbb{C}$ preserves the Hodge filtration (but is only semilinear with respect to τ for the action of $\mathbb{C}$). So all of these Hodge structures, maps, and extensions are actually in the category $\mathbb{Q} - \mathrm{MHS}^+$ of enriched mixed Hodge structures.

Let $V = H^0(W, \mathbb{Q})/\mathrm{im}(H^0(Y, \mathbb{Q}))$. Since $Y(\mathbb{C}) = \mathbb{C} \setminus \{1\}$, $H^1(Y, \mathbb{Q}) \simeq \mathbb{Q}$. As we will see there is a natural $\mathbb{Q}$-basis $c \in H^1(Y, \mathbb{Q})$, which we use to identify $H^1(Y, \mathbb{Q})$ with the 1-dimensional pure Hodge structure $\mathbb{Q}(-1)$ of weight 2 with ϕ_∞-action being multiplication by -1. Then the extension (4.2.b) together with the mixed Hodge structure on $H^1(Y, W, \mathbb{Q})$ and the action of ϕ_∞ defines a class $\mathcal{E}_{\mathrm{MH}} = [H^1(Y, W, \mathbb{Q})] \in \mathrm{Ext}^1_{\mathbb{Q}-\mathrm{MHS}^+}(\mathbb{Q}(-1), V)$. It is natural to ask:

$$\boxed{\text{is } \mathcal{E}_{\mathrm{MH}} \neq 0?}$$

That is, is (4.2.b) a nonsplit extension of enriched mixed Hodge structures?

Let $\lambda : V \twoheadrightarrow \mathbb{Q}(0)$ be any surjective map of (enriched) Hodge structures; since V is pure of weight 0, this is just any surjective linear map from V. Then the push-out of $\mathcal{E}_{\mathrm{MH}}$ by λ yields an extension $\mathcal{E}_{\mathrm{MH},\lambda} \in \mathrm{Ext}^1_{\mathbb{Q}-\mathrm{MHS}^+}(\mathbb{Q}(-1), \mathbb{Q}(0))$. It is clear that $\mathcal{E}_{\mathrm{MH}} \neq 0$ if and only if there exists some λ such that $\mathcal{E}_{\mathrm{MH},\lambda} \neq 0$. So it also natural, and even equivalent, to ask:

$$\boxed{\text{does there exists } \lambda \text{ such that } \mathcal{E}_{\mathrm{MH},\lambda} \neq 0?}$$

The keys to our answer to these question are

- explicit descriptions of some classes in $F^1H^1(Y, \mathbb{C})$ and $F^1H^1(Y, W, \mathbb{C})$ via their corresponding classes in $H^1_{\mathrm{dR}}(Y/\mathbb{C})$ and $H^1_{\mathrm{dR}}((Y, W)/\mathbb{C})$, and
- an analytic calculation with the explicit de Rham classes and their images under ϕ_∞.

These come together as follows: Let $0 \neq \omega \in H^0(\Omega^1_{X/\mathbb{C}}(\log Z))$. Via the de Rham isomorphism ι_{dR}, the differential ω determines classes $c = \iota_{\mathrm{dR}}([\omega]) \in F^1H^1(Y, \mathbb{C})$ and $c_H = \iota_{\mathrm{dR}}([\omega]) \in F^1H^1(Y, W, \mathbb{C})$ in the Hodge filtrations. Just as explained in the final paragraph of Section 4.1, $\mathcal{E}_{\mathrm{MH},\lambda} \neq 0$ if and only if the image of $(1+\phi_\infty)c_H \in V_{\mathbb{C}} = H^0(W, \mathbb{C})/\mathrm{im}(H^0(Y, \mathbb{C}))$ is nonzero under λ. In particular, $\mathcal{E}_{\mathrm{MH}} \neq 0$ if and only if $(1+\phi_\infty)c_H \neq 0$ in $V_{\mathbb{C}} = H^0(W, \mathbb{C})/\mathrm{im}(H^0(Y, \mathbb{C}))$. In some instances ω can be chosen so that $(1 + \phi_\infty)\omega = d\eta$ for some explicit real analytic function η on Y. Then $(1 + \phi_\infty)c_H$ is just the image of $\eta|_W \in H^0(W, \mathbb{C})$. In particular, $\mathcal{E} \neq 0$ if and only if $\lambda(\eta|_W) \neq 0$ for some homomorphism $\lambda : H^0(W, \mathbb{C}) \to \mathbb{C}$ (not necessarily $\mathbb{Q}$-valued) that is trivial on the image of $H^0(Y, \mathbb{C})$. In particular, to show that $\mathcal{E} \neq 0$

it will be enough to write down a sufficiently explicit ω so that η can be determined and seen to satisfy $\lambda(\eta|_W) \neq 0$ for some such λ. Note that $\lambda(\eta|_W)$ is just a linear combination of the values of η on the points in W.

Arguing this way we will show that $\mathcal{E}_{\mathrm{MH}} \neq 0$ if, for example, there exists a nontrivial even primitive Dirichlet character $\chi : (\mathbb{Z}/N\mathbb{Z})^\times \to \mathbb{C}^\times$. For a more precise result, see (4.6.a) below.

4.3. *The cohomology of Y.* As $Y(\mathbb{C})$ is just the Riemann sphere $X(\mathbb{C}) = \mathbb{P}^1(\mathbb{C})$ minus the two points ∞ and 1, the singular cohomology group $H^1(Y, \mathbb{Q})$ is isomorphic to $\mathbb{Q}$. A somewhat explicit isomorphism is given as follows.

Recall the long exact sequence for relative cohomology for the (open) inclusion $Y(\mathbb{C}) \subset X(\mathbb{C})$:

$$\cdots \to H^1(X, F) \to H^1(Y, F) \xrightarrow{\partial} H^2(X, Y, F) \\ \to H^2(X, F) \to H^2(Y, F) \to \cdots . \quad (4.3.a)$$

The group $H^2(X, Y, F)$ is naturally identified with the space $H^0(Z, F)(-1) = \bigoplus_{z\in Z(\mathbb{C})} (2\pi i)^{-1} F$. Under this identification, the induced map $H^0(Z, F) \to H^2(X, F)(1)$ is just the cycle class map.[5] In particular, Poincaré duality canonically identifies $H^2(X, F)(1)$ with the F-dual of $H^0(X, F)$ and the map $H^0(Z, F) \to H^2(X, F)(1)$ with the dual of the natural map $H^0(X, F) \to H^0(Z, F)$, $H^0(Z, F)$ being, of course, self-dual in the obvious way. As $H^1(X, F) = H^1(\mathbb{P}^1, F) = 0$, it follows that

$$\partial : H^1(Y, F) \xrightarrow{\sim} \Big\{((2\pi i)^{-1} a_z)_{z\in Z(\mathbb{C})} : a_z \in F,\ \sum_{z\in Z(\mathbb{C})} a_z = 0\Big\} \\ \subset H^0(Z, F)(-1). \quad (4.3.b)$$

In particular, as $Z(\mathbb{C}) = \{\infty, 1\}$,

$$\partial : H^1(Y, F) \xrightarrow{\sim} \big\{((2\pi i)^{-1} a, -(2\pi i)^{-1} a) : a \in F\big\} \simeq F.$$

As the Hodge structure on $H^0(Z, \mathbb{Q})$ is pure of weight 0, $H^0(Z, \mathbb{Q})(-1)$ is pure of weight 2, and so the isomorphism (4.3.b) implies that the Hodge structure on $H^1(Y, \mathbb{Q})$ is pure of weight 2. This can also be seen as follows. The Hodge filtration $F^\bullet H^1(Y, \mathbb{C})$ on $H^1(Y, \mathbb{C}) = H^1(Y, \mathbb{Q}) \otimes \mathbb{C}$ is defined via the de Rham isomorphism $\iota_{\mathrm{dR}} : H^1_{\mathrm{dR}}(Y/\mathbb{C}) \xrightarrow{\sim} H^1(Y, \mathbb{C})$ and the Hodge filtration on $H^1_{\mathrm{dR}}(Y/\mathbb{C})$. The de Rham cohomology $H^*_{\mathrm{dR}}(Y/\mathbb{C})$ is computed by the hypercohomology of both the de Rham complex $DR_Y = [\mathcal{O}_Y \xrightarrow{d} \Omega^1_Y]$ and the log de Rham complex $DR_X(\log Z) = [\mathcal{O}_X \xrightarrow{d} \Omega^1_X(\log Z)]$; the natural map $DR_X(\log Z) \to DR_Y$ is a quasi-isomorphism. Here '$(\log Z)$' denotes the complex with log poles along Z

[5] This is essentially the definition of the cycle class map.

(for more on the log de Rham complex and its cohomology, see [Kedlaya 2008, §1.9], [Esnault and Viehweg 1992, §2], or [Peters and Steenbrink 2008, §4]). Let $A = \mathbb{Q}\big[t, \frac{1}{t-1}\big]$. Since $Y = \operatorname{Spec} A$ is affine, the hypercohomology of the de Rham complex for Y is the cohomology of the complex itself. The Hodge filtration $F^\bullet H^1_{\mathrm{dR}}(Y/\mathbb{C})$ is just the image of the hypercohomology of the usual filtration on $DR_X(\log Z)$. In particular,

$$\begin{aligned} F^0 H^1_{\mathrm{dR}}(Y/\mathbb{C}) &= H^1_{\mathrm{dR}}(Y/\mathbb{C}) = \Omega^1_{A\otimes\mathbb{C}}/d(A\otimes\mathbb{C}) = \mathbb{C}\frac{dt}{1-t}, \\ F^1 H^1_{\mathrm{dR}}(Y/\mathbb{C}) &= \operatorname{im}(H^0(\Omega^1_X(\log Z))) = H^0(\Omega^1_X(\log Z)) = \mathbb{C}\frac{dt}{1-t}, \\ F^2 H^1_{\mathrm{dR}}(Y/\mathbb{C}) &= 0. \end{aligned}$$

The weight filtration $W_\bullet H^1(Y, \mathbb{Q})$ on $H^1(Y, \mathbb{Q})$ is given by $0 = W_0 H^1(Y, \mathbb{Q}) = W_1 H^1(Y, \mathbb{Q}) = \operatorname{im}(H^1(X, \mathbb{Q})) \subset W_2 H^1(Y, \mathbb{Q}) = H^1(Y, \mathbb{Q})$. Indeed, in this case the n-th part $W_n H^1(Y, \mathbb{C})$ of the weight filtration is the image of the hypercohomology of $W_{n-1} DR_X(\log Z)$, where $W_n DR_X(\log Z) = [0]$ $(n < 0)$, $W_0 DR_X(\log Z) = DR_X$, and $W_m DR_X(\log Z) = DR_X(\log Z)$ $(m \geq 1)$; see [Peters and Steenbrink 2008, Theorem 4.2]. It is a fundamental result of Hodge theory that this weight filtration on $H^1(Y, \mathbb{C})$ is actually rational.

Note that the compatibility of the de Rham isomorphisms with the long exact sequence (4.3.a) shows that the class $\iota_{\mathrm{dR}}([\omega_a]) \in H^1(Y, \mathbb{C})$ of the differential $\omega_a = (2\pi i)^{-1} a \frac{dt}{1-t} \in H^0(\Omega^1_{X/\mathbb{C}}(\log Z))$ satisfies

$$\partial(\iota_{\mathrm{dR}}([\omega]) = ((2\pi i)^{-1} a_\infty, (2\pi i)^{-1} a_1), \quad a_\infty = -a_1 = a. \tag{4.3.c}$$

This is because the corresponding boundary map for de Rham cohomology just takes the class of ω to $(\operatorname{Res}_z(\omega))_{z\in Z}$. In particular, the de Rham isomorphism induces an identification

$$H^1_{\mathrm{dR}}(Y/\mathbb{C}) \supset (2\pi i)^{-1} F \frac{dt}{1-t} \xrightarrow[\sim]{\iota_{\mathrm{dR}}} H^1(Y, F) \tag{4.3.d}$$

for any subfield $F \subset \mathbb{C}$.

Let

$$\omega = \frac{dt}{1-t} \in H^0(\Omega^1_X(\log Z)) \quad \text{and} \quad \omega^{\mathrm{an}} = (2\pi i)^{-1}\omega.$$

Let $[\omega^{\mathrm{an}}] \in F^1 H^1_{\mathrm{dR}}(Y/\mathbb{C})$ be the corresponding class. Then

$$H^1_{\mathrm{dR}}(Y/\mathbb{C}) = F^1 H^1_{\mathrm{dR}}(Y/\mathbb{C}) = \mathbb{C}[\omega^{\mathrm{an}}].$$

Let

$$c = \iota_{\mathrm{dR}}([\omega^{\mathrm{an}}]) \in F^1 H^1(Y, \mathbb{C}).$$

Then $H^1(Y, \mathbb{C}) = F^1 H^1(Y, \mathbb{C}) = \mathbb{C}c$. It follows from (4.3.d) that $c \in H^1(Y, \mathbb{Q})$.

4.4. ***The cohomology of*** (Y, W). We can compute the relative singular cohomology groups $H^1(Y, W, F)$ as the cohomology of the mapping cone $\mathrm{Cone}(C^\bullet(Y, F) \to C^\bullet(W, F))[-1]$ for $C^\bullet(Y, F)$, $C^\bullet(W, F)$ the singular cochain complexes with F coefficients and the map being that induced by the inclusion $W(\mathbb{C}) \hookrightarrow Y(\mathbb{C})$. Concretely, this mapping cone is $C^\bullet(Y, F) \oplus C^{\bullet-1}(W, F))$ with differential $d(a, b) = (d^\bullet a, -d^{\bullet-1} b - a|_W)$; the map to $C^\bullet(Y)$ is just projection onto the first summand.

The de Rham cohomology of the pair is similarly computed but with $C^\bullet(Y, F)$ and $C^\bullet(W, F)$ replaced by the de Rham complexes DR_Y and DR_W, respectively, or even $DR_X(\log Z)$ and DR_W. From the definition of the mapping cone, it is easy to see that $\mathrm{Cone}(DR_Y \to DR_W)[-1]$ (resp. $\mathrm{Cone}(DR_X(\log Z) \to DR_W)[-1]$) can be replaced with the quasiisomorphic subcomplex $DR_Y(-W) = [\mathcal{O}_Y(-W) \xrightarrow{d} \Omega^1_Y]$ (resp. $DR_X(\log Z)(-W) = [\mathcal{O}_X(-W) \xrightarrow{d} \Omega^1_X(\log Z)]$).

The Hodge filtration on $H^1(Y, W, \mathbb{C})$ is again defined via the de Rham isomorphism. In particular, it is given by the images of the hypercohomology of the usual filtration on $DR_X(\log Z)(-W)$. Recall that $A = \mathbb{Q}\big[t, \frac{1}{t-1}\big]$. Much as for $H^1_{\mathrm{dR}}(Y, \mathbb{C})$, we have

$$F^0 H^1_{\mathrm{dR}}((Y, W)/\mathbb{C}) = H^1_{\mathrm{dR}}((Y, W)/\mathbb{C}) = \Omega^1_{A\otimes\mathbb{C}}/d(\Phi_N(t) A \otimes \mathbb{C}),$$

$$F^1 H^1_{\mathrm{dR}}((Y, W)/\mathbb{C}) = \mathrm{im}(H^0(\Omega^1_X(\log Z))) = H^0(\Omega^1_X(\log Z)) = \mathbb{C}\frac{dt}{t-1},$$

$$F^2 H^1_{\mathrm{dR}}((Y, W)/\mathbb{C}) = 0.$$

The weight filtration on $H^1(Y, W, \mathbb{Q})$ is $W_\bullet H^1(Y, W, \mathbb{Q})$ with $W_{-1} H^1(Y, W, \mathbb{Q}) = 0$, $W_0 H^1(Y, W, \mathbb{Q}) = W_1 H^1(Y, W, \mathbb{Q}) = \mathrm{im}(H^0(W, \mathbb{Q}))$, and $W_2 H^1(Y, W, \mathbb{Q}) = H^1(Y, W, \mathbb{Q})$. Note that $W_0 H^1(Y, W, \mathbb{Q})/W_{-1} H^1(Y, W, \mathbb{Q}) = \mathrm{im}(H^0(W, \mathbb{Q}))$ and the induced Hodge filtration is indeed the unique Hodge structure pure of weight 0. (For more on the Hodge structures on relative cohomology see [Peters and Steenbrink 2008, §5.5].) Note also that $W_2 H^1(Y, W, \mathbb{Q})/W_1 H^1(Y, W, \mathbb{Q}) \xrightarrow{\sim} H^1(Y, \mathbb{Q})$ and the induced Hodge filtration is just the one on $H^1(Y, \mathbb{Q})$ described above. This just makes explicit in this setting the general fact that the extension (4.2.b) realizes the mixed Hodge structure on $H^1(Y, W, \mathbb{Q})$ as an extension in the category of mixed Hodge structures.

For ω^{an} as before, let $[\omega^{\mathrm{an}}]_W \in F^1 H^1_{\mathrm{dR}}((Y, W)/\mathbb{C})$. Then $F^1 H^1(Y, W, \mathbb{C}) = \mathbb{C}\iota_{\mathrm{dR}}([\omega^{\mathrm{an}}]_W)$, and the isomorphism $F^1 H^1(Y, W, \mathbb{C}) \xrightarrow{\sim} F^1 H^1(Y, \mathbb{C})$ maps $c_H = \iota_{\mathrm{dR}}([\omega^{\mathrm{an}}]_W)$ to $c = \iota_{\mathrm{dR}}([\omega^{\mathrm{an}}])$.

4.5. ***The involution*** ϕ_∞. Since each of the varieties X, Y, Z, and W is defined over $\mathbb{Q}$, the cohomology groups considered above — singular and de Rham — all carry the action of an involution, denoted ϕ_∞, induced by the action of complex conjugation on the $\mathbb{C}$-points of the varieties. The maps in (4.3.a), (4.3.b), and (4.2.a) are all compatible with the actions of ϕ_∞ as are the de Rham isomorphisms ι_{dR}.

Moreover, ϕ_∞ interacts well with the Hodge filtrations: $\phi_\infty(F^p(-)) = \overline{F^p(-)}$. That is, $\phi_\infty \circ \tau$, for τ the action of complex conjugation on the coefficients, preserves the Hodge filtrations. Since $\overline{F^1H^1(Y,\mathbb{C})} = F^1H^1(Y,\mathbb{C})$ and $F^1H^1(Y,W,\mathbb{C}) \xrightarrow{\sim} F^1H^1(Y,\mathbb{C})$, the extension (4.2.b) is a split extension of Hodge structures enriched with the involutions ϕ_∞ if and only if $\overline{F^1H^1(Y,W,\mathbb{C})} = F^1H^1(Y,W,\mathbb{C})$. But this is often not the case, as we show below.

4.6. ***The calculation.*** The class $\phi_\infty(\iota_{\mathrm{dR}}([\omega^{\mathrm{an}}]_W)) = \iota_{\mathrm{dR}}(\phi_\infty([\omega^{\mathrm{an}}]_W))$ is represented by the real analytic differential $\phi_\infty^*\omega^{\mathrm{an}} = (2\pi i)^{-1}\frac{d\bar t}{1-\bar t}$ via the real-analytic de Rham isomorphism. Then $(1+\phi_\infty)\iota_{\mathrm{dR}}([\omega^{\mathrm{an}}]_W)$ is represented by the real analytic differential

$$(2\pi i)^{-1}\left(\frac{dt}{1-t} + \frac{d\bar t}{1-\bar t}\right) = -(2\pi i)^{-1} d\log|t-1|^2 = d(-(2\pi i)^{-1}\log|t-1|^2).$$

Let $\eta = -(2\pi i)^{-1}\log|t-1|^2$. This is a real-analytic function on Y. It follows that $(1+\phi_\infty)\iota_{\mathrm{dR}}([\omega^{\mathrm{an}}]_W)$ is the image of the class $\eta|_W \in H^0_{\mathrm{dR}}(W/\mathbb{C}) = H^0(W,\mathbb{C})$, which is just $(\eta(\zeta))_{\zeta\in W(\mathbb{C})} = \bigoplus_{\zeta\in W(\mathbb{C})}\mathbb{C} = H^0(W,\mathbb{C})$.

Let $\zeta_N = \exp(2\pi i/N) \in \mu_N^\circ(\mathbb{C}) = W(\mathbb{C})$. Let $\chi : (\mathbb{Z}/N\mathbb{Z})^\times \to \mathbb{C}^\times$ be a nontrivial character and let

$$\lambda_\chi : H^0(W,\mathbb{C}) \to \mathbb{C}, \quad \lambda_\chi((x_\zeta)_{\zeta\in W(\mathbb{C})}) = \sum_{a\in(\mathbb{Z}/N\mathbb{Z})^\times} \chi(a)x_{\zeta_N^a}.$$

As χ is nontrivial, λ_χ is 0 on the image of $H^0(Y,\mathbb{C})$, which is just the image of the diagonal embedding $\mathbb{C} \hookrightarrow \bigoplus_{\zeta\in W(\mathbb{C})}\mathbb{C}$. Then

$$\lambda_\chi(\eta|_W) = -2(2\pi i)^{-1}\sum_{a\in(\mathbb{Z}/N\mathbb{Z})^\times}\chi(a)\log|\zeta_N^a - 1|.$$

If χ is odd (so $\chi(-1) = -1$) then the sum is 0 as $|\zeta_N^a - 1| = |\zeta_N^{-a} - 1|$. But if χ is even (so $\chi(a) = \chi(-a)$), then the sum equals

$$2(2\pi i)^{-1}\frac{N_0}{\tau(\bar\chi_0)}L(1,\bar\chi_0)\prod_{\substack{\ell \text{ prime}\\ \ell\mid N\\ \ell\nmid N_0}}(1-\chi_0(\ell))$$

by a well-known formula for the value of the Dirichlet series $L(s,\bar\chi_0)$ at the point $s=1$ (see [Washington 1997, Theorem 4.9]). Here χ_0 is the primitive character associated with χ, N_0 is its conductor, $\bar\chi_0 = \chi_0^{-1}$, and $\tau(\bar\chi_0)$ is the usual Gauss sum. By the functional equation for $L(s,\chi_0)$, the last displayed expression equals

$$-(2\pi i)^{-1}4L'(0,\chi_0)\prod_{\substack{\ell \text{ prime}\\ \ell\mid N\\ \ell\nmid N_0}}(1-\chi_0(\ell)) = -(2\pi i)^{-1}4L'(0,\chi).$$

As noted before, $\mathcal{E}_{\mathrm{MH}}$ is a nonsplit extension of enriched Hodge structures if and only if $\lambda(\eta|_W) \neq 0$ for some $\lambda : H^0(W, \mathbb{C}) \to \mathbb{C}$ that vanishes on the image of $H^0(Y, \mathbb{C})$. Such λ are exactly the linear combinations of the λ_χ for χ running over the nontrivial characters of $(\mathbb{Z}/N\mathbb{Z})^\times$. So as a consequence of the calculation above we have:

$$\boxed{\begin{array}{c}\text{there is a nontrivial even character } \chi : (\mathbb{Z}/N\mathbb{Z})^\times \to \mathbb{C}^\times \\ \text{such that } \operatorname{ord}_{s=0} L(s, \chi) = 1\end{array} \iff \mathcal{E}_{\mathrm{MH}} \neq 0.} \tag{4.6.a}$$

The left-hand side is satisfied, of course, if there is a primitive even character modulo N.

Suppose that χ is quadratic as well as even. Then $\mathcal{E}_{\mathrm{MH},\chi} = H^1(Y, W, \mathbb{Q})/\ker(\lambda_\chi)$ is an extension of enriched Hodge structures that fits into a commutative diagram:

$$\begin{tikzcd} \frac{H^0(W,\mathbb{Q})}{\operatorname{im}(H^0(Y,\mathbb{Q}))} \arrow[r, hook] \arrow[d, two heads, "\lambda_\chi"] & H^1(Y, W, \mathbb{Q}) \arrow[r, two heads] \arrow[d, two heads, "/\ker(\lambda_\chi)"] & H^1(Y, \mathbb{Q}) \arrow[d, equal] \\ \mathbb{Q} \arrow[r, hook] & \mathcal{E}_{\mathrm{MH},\chi} \arrow[r, two heads] & \mathbb{Q}c. \end{tikzcd}$$

As $\ker(\lambda_\chi)$ is clearly stable under ϕ_∞,

$$\mathcal{E}_{\mathrm{MH},\chi} \in \operatorname{Ext}^1_{\mathbb{Q}-\mathrm{MHS}^+}(\mathbb{Q}c, \mathbb{Q}) = \operatorname{Ext}^1_{\mathbb{Q}-\mathrm{MHS}^+}(\mathbb{Q}(-1), \mathbb{Q}(0)).$$

Note that $\mathcal{E}_{\mathrm{MH},\chi}$ is just the image of $\mathcal{E}_{\mathrm{MH}}$ under the map induced by λ_χ. The calculation above shows

$$\boxed{\chi \text{ even and nontrivial, } \operatorname{ord}_{s=0} L(s, \chi) = 1 \iff \mathcal{E}_{\mathrm{MH},\chi} \neq 0.} \tag{4.6.b}$$

4.6.1. Remark. The fact that $\mathcal{E}_{\mathrm{MH},\chi}$ is split when χ is odd is consistent with the fact that $L(0, \chi) \neq 0$ for χ odd and primitive, and so we do not expect extensions.

5. Nonsplit extensions of *p*-adic Galois representations

We explain how arguments analogous to those in Section 4 yield statements analogous to (4.6.a) and (4.6.b) for certain extensions of p-adic Galois representations that occur in the relative étale cohomology of the pair (Y, W). Just as in the case of the extensions of mixed Hodge structures, we check that these Galois extensions are nonsplit by integrating an explicit differential representing a class in the cohomology of Y and recognizing the resulting formulas as expressions for L-values of Dirichlet characters. Only in this case the integration takes place in the context of p-adic rigid analysis, and the passage from étale cohomology to rigid geometry goes via the comparison theorems of p-adic Hodge theory. We explain how this calculation essentially computes the Bloch–Kato logarithm of these Galois extensions. We also explain how these extensions are the first layer of an Euler system. A reader with some familiarity with p-adic Hodge theory should be able to grasp the gist quickly upon reading Section 5.2 and fill in details by scanning the subsequent displayed equations. For readers less familiar with p-adic Hodge theory, we have included — much as we did in Section 4 — a brief introduction and some useful resources, particularly in Sections 5.1, 5.3, and 5.4.

Conventions. Let $\overline{\mathbb{Q}}$ be a fixed separable algebraic closure of $\mathbb{Q}$. We fix an embedding $\iota_\infty : \overline{\mathbb{Q}} \hookrightarrow \mathbb{C}$, which we use to identify $\overline{\mathbb{Q}}$ as a subfield of $\mathbb{C}$. For each prime ℓ we also fix a separable algebraic closure $\overline{\mathbb{Q}}_\ell$ of $\mathbb{Q}_\ell$ and an embedding $\iota_\ell : \overline{\mathbb{Q}} \hookrightarrow \overline{\mathbb{Q}}_\ell$. The latter identifies $G_{\mathbb{Q}_\ell} = \mathrm{Gal}(\overline{\mathbb{Q}}_\ell/\mathbb{Q}_\ell)$ with a decomposition group of $G_\mathbb{Q}$ for the prime ℓ. We let $I_\ell \subset G_{\mathbb{Q}_\ell}$ be the inertia subgroup and $\mathrm{frob}_\ell \in G_{\mathbb{Q}_\ell}/I_\ell$ the arithmetic Frobenius element. In particular, we identify $\overline{\mathbb{Q}}$ with a subfield of $\overline{\mathbb{Q}}_p$ via ι_p. Let $\mathbb{Q}_p^{\mathrm{ur}} \subset \overline{\mathbb{Q}}_p$ be the maximal unramified extension of $\mathbb{Q}_p$.

Let $\epsilon : G_\mathbb{Q} \to \mathbb{Z}_p^\times$ be the p-adic Galois character giving the action of $G_\mathbb{Q}$ on all p-th-power roots of unity and so on $\mathbb{Z}_p(1) = \varprojlim_r \mu_{p^r}$. The exponential map $\exp : \mathbb{Z}(1) \to \mathbb{C}^\times$ identifies $\varprojlim_r(\mathbb{Z}(1) \otimes \mathbb{Z}/p^r\mathbb{Z})$ with $\varprojlim_r \mu_{p^r} = \mathbb{Z}_p(1)$. We let $\underline{\zeta} \in \mathbb{Z}_p(1)$ be the $\mathbb{Z}_p$-basis that is the image of $2\pi i \in \mathbb{Z}(1)$.

Given a variety $\mathcal{V}$ defined over $\mathbb{Q}$ we let $\overline{\mathcal{V}}$ denote its base change $\mathcal{V}_{/\overline{\mathbb{Q}}}$ over $\overline{\mathbb{Q}}$. The role of the de Rham-singular isomorphisms in the preceding section will here be played by the de Rham-étale comparison isomorphisms of p-adic Hodge theory. This essentially allows us to compare $H^1_{\text{ét}}(\overline{\mathcal{V}}, \mathbb{Q}_p)$ with $H^1_{\mathrm{dR}}(\mathcal{V}/\mathbb{Q}_p)$ (for good $\mathcal{V}$), with the additional complication that the comparison is not direct but passes through the D_{dR}-functor: for a finite-dimensional continuous $\mathbb{Q}_p$-linear $G_{\mathbb{Q}_p}$-representation M, $D_{\mathrm{dR}}(M) = (M \otimes_{\mathbb{Q}_p} B_{\mathrm{dR}})^{G_{\mathbb{Q}_p}}$, where B_{dR} is the usual de Rham period ring. The de Rham-étale comparison isomorphism is a canonical functorial isomorphism $\iota_{\mathrm{dR},p}$ of $H^1_{\mathrm{dR}}(\mathcal{V}/\mathbb{Q}_p)$ with $D_{\mathrm{dR}}(H^1_{\text{ét}}(\overline{\mathcal{V}}, \mathbb{Q}_p))$, and similarly for relative cohomology with respect to a subvariety $\mathcal{U} \subset \mathcal{V}$ (at least for $\mathcal{U}$ a normal crossings divisor or a complement of such). These isomorphisms are functorial in $\mathcal{V}$ and compatible with the long-exact sequences for relative cohomology, Gysin sequences, etc.

5.1. ***p-adic Galois representations and their period rings.*** Let $L/\mathbb{Q}_p$ be a finite extension of $\mathbb{Q}_p$ and let V be a finite-dimensional L-space equipped with a continuous L-linear action of $G_{\mathbb{Q}_p}$. Simple examples of such are the one-dimensional $\mathbb{Q}_p$-spaces $\mathbb{Q}_p(n) = (\mathbb{Z}_p(1)^{\otimes n}) \otimes_{\mathbb{Z}_p} \mathbb{Q}_p$, on which $G_{\mathbb{Q}_p}$ acts[6] via the n-th-power ϵ^n of the p-adic cyclotomic character ϵ. For any V we write $V(n)$ for $V \otimes_{\mathbb{Q}_p} \mathbb{Q}_p(n)$ (the n-th Tate twist of V) with $G_{\mathbb{Q}_p}$ acting on both factors. If $\chi : G_{\mathbb{Q}_p} \to L^\times$ is any continuous character, then we let $L(\chi)$ be the one-dimensional L-space with $\sigma \in G_{\mathbb{Q}_p}$ acting as multiplication by $\chi(\sigma)$. Unlike for $\mathbb{Q}_p(n)$, the representation

[6]Of course, the Galois group $G_\mathbb{Q}$ also acts on $\mathbb{Q}_p(n)$. In fact, in subsequent sections we will largely be interested in $G_{\mathbb{Q}_p}$-actions that are the restrictions of $G_\mathbb{Q}$-actions.

$L(\chi)$ has an implicit L-basis; hence, identifying $L(\chi)(n)$ with $L(\chi\epsilon^n)$ requires choosing a basis of $\mathbb{Q}_p(n)$. Other important examples of V's arise in arithmetic geometry. For a complete, geometrically connected variety $\mathcal{V}$ of dimension d defined over a subfield of $\mathbb{Q}_p$, $H^{2d}_{\text{ét}}(\mathcal{V}_{/\overline{\mathbb{Q}}_p}, \mathbb{Q}_p)$ is isomorphic to $\mathbb{Q}_p(-d)$. More generally, if $\mathcal{V}$ is any variety over $\mathbb{Q}_p$, then the étale cohomology groups $H^*_{\text{ét}}(\mathcal{V}_{/\overline{\mathbb{Q}}_p}, \mathbb{Q}_p)$ (as well as relative cohomology groups) are finite-dimensional $\mathbb{Q}_p$-spaces with $\mathbb{Q}_p$-linear continuous actions of $G_{\mathbb{Q}_p}$, and all the maps in the associated exact sequences (e.g., the Gysin sequences and the sequences for relative cohomology) are maps of such representations.

There are subclasses of p-adic Galois representations that figure prominently in arithmetic geometry:

$$\begin{Bmatrix}\text{crystalline}\\ \text{reps.}\end{Bmatrix} \subset \begin{Bmatrix}\text{semistable}\\ \text{reps.}\end{Bmatrix} \subset \begin{Bmatrix}\text{potentially}\\ \text{semistable}\\ \text{reps.}\end{Bmatrix} = \begin{Bmatrix}\text{de Rham}\\ \text{reps.}\end{Bmatrix}.$$

Each class is characterized by a period ring $B_?$, $? = \text{crys, st, or dR}$, respectively. These period rings are topological $\mathbb{Q}_p^{\text{ur}}$-algebras (even domains), and even a $\overline{\mathbb{Q}}_p$-algebra in the case of B_{dR}. Each is equipped with a continuous action of $G_{\mathbb{Q}_p}$ compatible with the action on $\mathbb{Q}_p^{\text{ur}}$ (on $\overline{\mathbb{Q}}_p$ in the case of B_{dR}) and such that $B_?^{G_{\mathbb{Q}_p}} = \mathbb{Q}_p$. It is always true that the $\mathbb{Q}_p$-dimension of $D_?(V) := (V \otimes_{\mathbb{Q}_p} B_?)^{G_{\mathbb{Q}_p}}$ is at most that of V, and by definition V belongs to the corresponding class for ? if and only if the $\mathbb{Q}_p$-dimension of $D_?(V)$ equals the $\mathbb{Q}_p$-dimension of V. (If $G_{\mathbb{Q}_p}$ were replaced by G_K for a general finite extension $K/\mathbb{Q}_p$, the picture would be slightly different.) If V is an L-space, then so is $D_?(V)$ and one can also check whether V belongs to the category ? by comparing dimensions over L. The ring B_{crys} is a subring of both B_{dR} and B_{st}, and if we fix a branch of the p-adic logarithm, then B_{st} can viewed as a sub-B_{crys}-algebra of B_{dR}. There is a canonical inclusion $\mathbb{Z}_p(1) \hookrightarrow B_{\text{crys}}$ and we let $\underline{t} \in B_{\text{crys}}$ be the image of $\underline{\zeta}$. The element $\underline{t}$ is invertible in B_{crys} (hence also in the other rings) and

$$D_?(\mathbb{Q}_p(n)) = \mathbb{Q}_p(\underline{\zeta}^{\otimes n} \otimes \underline{t}^{-n}).$$

So the representations $\mathbb{Q}_p(n)$ are all crystalline. Clearly then, V is crystalline (or semistable or de Rham) if and only if $V(n)$ is for some integer n. More generally, each of these classes of representations is stable under direct sums, duals, tensor products, taking subrepresentations or quotients (and hence subquotients). However, they are not closed under extensions. For more on p-adic Galois representations and these period rings the interested reader should consult [Berger 2004; 2013] or [Conrad and Brinon 2009].

Suppose $\mathcal{V}$ is a variety over $\mathbb{Q}_p$. The étale cohomology groups $H^*_{\text{ét}}(\mathcal{V}_{/\overline{\mathbb{Q}}_p}, \mathbb{Q}_p)$ are all de Rham (equivalently, potentially semistable), as are the relative cohomology groups. If $\mathcal{V}$ has a smooth complete model over $\mathbb{Z}_p$, then $H^*_{\text{ét}}(\mathcal{V}_{/\overline{\mathbb{Q}}_p}, \mathbb{Q}_p)$ is crystalline. Not at all surprisingly, if $\mathcal{V}$ has a semistable model over $\mathbb{Z}_p$, then $H^*_{\text{ét}}(\mathcal{V}_{/\overline{\mathbb{Q}}_p}, \mathbb{Q}_p)$ is semistable.

The ring B_{dR} of de Rham periods has a natural decreasing filtration: $F^i B_{\text{dR}} = \underline{t}^i B^+_{\text{dR}}$ for a certain subring B^+_{dR} containing $\underline{t}$. This induces a finite exhaustive filtration on $D_{\text{dR}}(V)$ for any V: $F^i D_{\text{dR}}(V) = (V \otimes_{\mathbb{Q}_p} \underline{t}^i B^+_{\text{dR}})^{G_{\mathbb{Q}_p}}$. The rings B_{crys} and B_{st} are equipped with a semilinear Frobenius ϕ_p. That is, ϕ_p is a continuous endomorphism that acts semilinearly with respect to the usual (arithmetic) Frobenius frob_p on the maximal unramified extension $\mathbb{Q}_p^{\text{ur}}$ of $\mathbb{Q}_p$ (so $\phi(ax) = \text{frob}_p(a)\phi(x)$ for $a \in \mathbb{Q}_p^{\text{ur}}$ and $x \in B_{\text{crys}}, B_{\text{st}}$). The ring B_{st} also has a nilpotent endomorphism N (sometimes called a monodromy operator) such satisfying $N\phi_p = p\phi_p N$. The Frobenius ϕ_p acts on $\underline{t}$ as multiplication by p. In particular, in the case of $\mathbb{Q}_p(n)$ we have $D(n) := D_{\text{dR}}(\mathbb{Q}_p(n)) = D_{\text{crys}}(\mathbb{Q}_p(n)) = \mathbb{Q}_p(\underline{\zeta}^{\otimes n} \otimes \underline{t}^{-n})$ is a free $\mathbb{Q}_p$-space of rank one with ϕ_p acting as multiplication by p^{-n}. The filtration on $D(n)$ is such that $F^i D(n) = D(n)$ if $i \leq -n$ and $F^i D = 0$ otherwise.

Analogously to Section 4, in the rest of Section 5 we will construct an extension of $\mathbb{Q}_p(-1)$ by some $\mathbb{Q}_p(\chi)$, for χ a finite character, in the category of crystalline representations of $G_{\mathbb{Q}_p}$. We will investigate when this extension is nonsplit. Suppose then that we have a crystalline extension

$$0 \to V \to E \to \mathbb{Q}_p(m) \to 0.$$

Applying the $D_{\text{crys}}(-)$ functor we obtain an extension

$$0 \to D(V) \to D(E) \to D(m) \to 0$$

of filtered $\mathbb{Q}_p$-spaces with a $\mathbb{Q}_p$-linear action of ϕ_p. The extension E is split if and only if the extension $D(E)$ is.[7] Suppose that $F^{-m} D(V) = 0$. Let $0 \neq \omega \in D(m)$ and $\omega_H \in F^{-m} D(E)$ that maps to 0. As $F^{-m} D(E) \cap D(V) = 0$, the $\mathbb{Q}_p$-map $\phi : D(m) \to D(E)$ that takes ω to ω_H is the unique splitting of E as filtered $\mathbb{Q}_p$-spaces. It follows that the extension $D(E)$ is split if and only if $\phi(\phi_p \omega) = \phi_p \phi(\omega) = \phi_p \omega_H$. As $\phi_p \omega = p^m \omega$, this holds if and only if $\phi_p \omega_H = p^m \omega_H$, or, equivalently, $(1 - p^{-m}\phi_p)\omega_H = 0$.

In practice, we will be able to use Hodge theory to find ω_H and to compute $(1 - p^{-m}\phi_p)\omega_H$.

5.2. *The extension $\mathcal{E}_{\mathbb{Q}_p,\text{ét}}$.* Let $F/\mathbb{Q}_p$ be any finite extension. From the long exact sequence

$$\cdots \to H^0_{\text{ét}}(\overline{Y}, F) \to H^0_{\text{ét}}(\overline{W}, F) \to H^1_{\text{ét}}(\overline{Y}, \overline{W}, F) \to H^1_{\text{ét}}(\overline{Y}, F) \to H^1_{\text{ét}}(\overline{W}, F) \to \cdots \tag{5.2.a}$$

[7] The ‘only if’ direction is clear. The ‘if’ direction is a consequence of the equivalence of crystalline representations and *admissible* filtered ϕ_p-modules.

of étale cohomology groups we obtain an extension of $F[G_{\mathbb{Q}}]$-modules

$$0 \to \frac{H^0_{\text{ét}}(\overline{W}, F)}{\text{im}(H^0_{\text{ét}}(\overline{Y}, F))} \to H^1_{\text{ét}}(\overline{Y}, \overline{W}, F) \to H^1_{\text{ét}}(\overline{Y}, F) \to 0. \tag{5.2.b}$$

Let $V_F = H^0_{\text{ét}}(\overline{W}, F)/\text{im}(H^0_{\text{ét}}(\overline{Y}, F))$. As we will see, $H^1_{\text{ét}}(\overline{Y}, F) \simeq F(-1)$ as $F[G_{\mathbb{Q}}]$-modules, and there is a natural F-basis $c \in H^1_{\text{ét}}(\overline{Y}, F)$, which we will use to identify $H^1_{\text{ét}}(\overline{Y}, F)$ with $F(-1)$. Then the extension (5.2.b) yields a class $\mathcal{E}_{F,\text{ét}} = [H^1_{\text{ét}}(\overline{Y}, \overline{W}, F)] \in \text{Ext}^1_{F[G_{\mathbb{Q}}]}(V_F, F(-1))$. This is just the p-adic étale analog of the extension class $\mathcal{E}$ of rational Hodge structures considered in the preceding section. As in that case, it is natural to ask:

$$\text{is } \mathcal{E}_{F,\text{ét}} \neq 0?$$

And much as before, the keys to our answer to this question are

- explicit descriptions of some classes in $H^1_{\text{ét}}(\overline{Y}, F)$ and the action of $G_{\mathbb{Q}}$ on these classes,
- the action of a p-adic Frobenius ϕ_p on the de Rham versions of the cohomology groups in (4.2.b) and its action on the de Rham realizations of the explicit classes, and
- the reduction, via p-adic Hodge theory, to a p-adic analytic calculation with the de Rham realizations of the explicit classes and their images under ϕ_p.

These combine to provide an answer to the question about the nonvanishing of $\mathcal{E}_{F,\text{ét}}$ much in the same way that their real and complex analogs answered the question about the nonvanishing of $\mathcal{E}_{\text{MH}}$.

5.3. ***The étale cohomology of $\overline{Y}$.*** In the long exact sequence for the relative étale cohomology for the (open) inclusion $Y \subset X$,

$$\begin{aligned}\cdots \to H^1_{\text{ét}}(\overline{X}, F) \to H^1_{\text{ét}}(\overline{Y}, F) \xrightarrow{\partial_{\text{ét}}} H^2_{\text{ét}}(\overline{X}, \overline{Y}, F) \\ \to H^2_{\text{ét}}(\overline{X}, F) \to H^2_{\text{ét}}(Y, F) \to \cdots,\end{aligned} \tag{5.3.a}$$

the group $H^2_{\text{ét}}(\overline{X}, \overline{Y}, F)$ is naturally identified with the space $H^0_{\text{ét}}(\bar{Z}, F(-1)) = \bigoplus_{z \in \underline{Z}(\overline{\mathbb{Q}})} F \otimes \underline{\zeta}^\vee$. This identification is such that the induced map $H^0_{\text{ét}}(\bar{Z}, F) \to H^2_{\text{ét}}(\overline{X}, F(1)) = F$ is just the cycle class map. It follows that

$$\begin{aligned}\partial_{\text{ét}} : H^1_{\text{ét}}(\overline{Y}, F) \xrightarrow{\sim} \Big\{ (a_z \otimes \underline{\zeta}^\vee)_{z \in Z(\overline{\mathbb{Q}})} : a_z \in F, \sum_{z \in Z(\overline{\mathbb{Q}})} a_z = 0 \Big\} \\ \subset H^0_{\text{ét}}(\bar{Z}, F(-1)).\end{aligned} \tag{5.3.b}$$

In particular, as $Z(\overline{\mathbb{Q}}) = \{\infty, 1\}$, $\partial_{\text{ét}} : H^1_{\text{ét}}(\overline{Y}, F) \xrightarrow{\sim} \{(a \otimes \underline{\zeta}^\vee, -a \otimes \underline{\zeta}^\vee) : a \in F\} \simeq F$. The action of $G_{\mathbb{Q}}$ on $H^1_{\text{ét}}(\overline{Y}, F)$ is easily read off from this: The Galois action on

$H^0_{\text{ét}}(\bar{Z}, F(-1))$ is just given by $\sigma(a)_z = \epsilon(\sigma)^{-1} a_{\sigma^{-1}(z)} \otimes \underline{\zeta}^\vee$ for $a = (a_z \otimes \underline{\zeta}^\vee)_{z \in Z(\overline{\mathbb{Q}})}$ and $\sigma \in G_{\mathbb{Q}}$. Since the points of Z are defined over $\mathbb{Q}$, this shows that $H^1_{\text{ét}}(\overline{Y}, F) \simeq F(-1)$ as an $F[G_{\mathbb{Q}}]$-module.

Let $c_{\text{ét}} \in H^1_{\text{ét}}(\overline{Y}, \mathbb{Q}_p)$ be the class corresponding under $\partial_{\text{ét}}$ to

$$(c_\infty, c_1) = (1 \otimes \underline{\zeta}^\vee, -1 \otimes \underline{\zeta}^\vee).$$

Then

$$\sigma c_{\text{ét}} = \epsilon^{-1}(\sigma) c_{\text{ét}}, \quad \sigma \in G_{\mathbb{Q}}, \tag{5.3.c}$$

and $H^1_{\text{ét}}(\overline{Y}, \mathbb{Q}_p) = \mathbb{Q}_p c_{\text{ét}} \simeq \mathbb{Q}_p(-1)$.

The singular-étale comparison isomorphisms $\iota_{\text{ét}}$ identify the sequence (4.2.a) with (5.2.a) and the isomorphism (4.3.b) with (5.3.b) (with $(2\pi i)^{-1}$ being identified with $1 \otimes \underline{\zeta}^\vee$). It follows that

$$\iota_{\text{ét}}(c) = c_{\text{ét}}.$$

However, this is not needed in the following.

5.4. *D_{dR} and the Frobenius ϕ_p.* The étale cohomology groups in (5.2.a) are all de Rham representations of $G_{\mathbb{Q}_p}$. In particular, applying the D_{dR}-functor yields a commutative diagram

$$\begin{tikzcd}
\frac{H^0_{\mathrm{dR}}(W/\mathbb{Q}_p)}{\mathrm{im}(H^0_{\mathrm{dR}}(Y/\mathbb{Q}_p))} \arrow[r, hook] \arrow[d, equal, "\iota_{\mathrm{dR},p}"] & H^1_{\mathrm{dR}}((Y,W)/\mathbb{Q}_p) \arrow[r, two heads, "\alpha_{\mathrm{dR}}"] \arrow[d, equal, "\iota_{\mathrm{dR},p}"] & H^1_{\mathrm{dR}}(Y/\mathbb{Q}_p) \arrow[d, equal, "\iota_{\mathrm{dR},p}"] \\
\frac{D_{\mathrm{dR}}(H^0_{\text{ét}}(\overline{W},\mathbb{Q}_p))}{\mathrm{im}(D_{\mathrm{dR}}(H^0_{\text{ét}}(\overline{Y},\mathbb{Q}_p))} \arrow[r, hook] & D_{\mathrm{dR}}(H^1_{\text{ét}}(\overline{Y},\overline{W},\mathbb{Q}_p)) \arrow[r, two heads, "\alpha_{\text{ét}}"] & D_{\mathrm{dR}}(H^1_{\text{ét}}(\overline{Y},\mathbb{Q}_p)),
\end{tikzcd} \tag{5.4.a}$$

where the vertical arrows are the de Rham comparison isomorphisms of p-adic Hodge theory. These spaces are all filtered $\mathbb{Q}_p$-spaces: the Hodge filtration on the top line is identified with the filtration induced from the filtration $\underline{t}^i B^+_{\mathrm{dR}}$ on B_{dR} on the bottom line. All the maps are morphisms of filtered $\mathbb{Q}_p$-spaces.

A splitting of the extension (5.2.b) would give a splitting of the bottom line of this diagram, and hence a splitting of the top, as filtered $\mathbb{Q}_p$-spaces. We will show that this does not happen in general, at least if we also take into account the action of an additional operator on these spaces — a Frobenius operator ϕ_p (which replaces ϕ_∞ in this p-adic context). The splittings would also be splittings for the action of ϕ_p, and we will show that such splittings do not exist when certain values of p-adic L-functions are nonzero.

To explain what ϕ_p is and illustrate its role, we make the simplifying hypothesis that

$$p \nmid N. \tag{5.4.b}$$

The varieties X, Y, and Z have smooth models over $\mathbb{Z}$ — just replace $\mathbb{Q}$ with $\mathbb{Z}$ — and W has a smooth model over $\mathbb{Z}[\frac{1}{N}]$ — just replace $\mathbb{Q}$ with $\mathbb{Z}[\frac{1}{N}]$. Hence they also all have smooth models $\mathscr{X}$, $\mathscr{Y}$, $\mathscr{Z}$, and $\mathscr{W}$ over $\mathbb{Z}_p$ under (5.4.b). The inclusions $W \hookrightarrow Y \hookrightarrow X$ and $Z \hookrightarrow X$ extend to these models. This implies that the cohomology groups in (5.2.b) are all crystalline representations of $G_{\mathbb{Q}_p}$, and so D_{dR} can be replaced with the crystalline functor D_{crys} in the bottom line of (5.4.a). The modules $D_{\mathrm{crys}}(-) = (- \otimes_{\mathbb{Q}_p} B_{\mathrm{crys}})^{G_{\mathbb{Q}_p}}$ inherit a $\mathbb{Q}_p$-linear action of the crystalline Frobenius ϕ_p from B_{crys}. In particular, if $\mathcal{E}_{\mathbb{Q}_p,\text{ét}}$ were 0 after restriction to $G_{\mathbb{Q}_p}$ then the bottom line in (5.4.a) would be simultaneously split as an extension of filtered $\mathbb{Q}_p$-spaces and as an extension of $\mathbb{Q}_p[\phi_p]$-modules.

From the Galois action (5.3.c) we see that

$$c_{\mathrm{crys}} = c_{\text{ét}} \otimes \underline{t} \in D_{\mathrm{crys}}(H^1(\overline{Y}, \mathbb{Q}_p)) = D_{\mathrm{dR}}(H^1(\overline{Y}, \mathbb{Q}_p))$$

is a $\mathbb{Q}_p$-basis of $D_{\mathrm{crys}}(H^1(\overline{Y}, \mathbb{Q}_p)) = D_{\mathrm{crys}}(\mathbb{Q}_p c)$. As ϕ_p acts on $\underline{t}$ as multiplication by p, it follows that

$$\phi_p c_{\mathrm{crys}} = p c_{\mathrm{crys}}. \tag{5.4.c}$$

Noting that $\omega = \frac{dt}{1-t} \in H^0(\Omega^1_{X/\mathbb{Q}_p}(\log Z))$, we let

$$c_{\mathrm{dR}} = [\omega] \in F^1 H^1_{\mathrm{dR}}(Y/\mathbb{Q}_p) \quad \text{and} \quad c_{\mathrm{dR},H} = [\omega]_W \in F^1 H^1_{\mathrm{dR}}((Y, W)/\mathbb{Q}_p).$$

The de Rham comparison isomorphisms of p-adic Hodge theory are compatible with the boundary map in the sequence (5.3.a), in the sense that

$$H^1_{\mathrm{dR}}(Y/\mathbb{Q}_p) \overset{\iota_{\mathrm{dR},p}}{=} D_{\mathrm{dR}}(H^1_{\text{ét}}(\overline{Y}, \mathbb{Q}_p)) \xrightarrow{\partial_{\text{ét}} \otimes \mathrm{id}} D_{\mathrm{dR}}(H^0_{\text{ét}}(\overline{Z}, \mathbb{Q}_p(-1))) \overset{\iota_{\mathrm{dR},p}}{=} H^0_{\mathrm{dR}}(Z/\mathbb{Q}_p)(-1)$$

is just the boundary map (the residue map) in the corresponding sequence for de Rham cohomology. As $\underline{\zeta}^\vee \otimes \underline{t}$ is identified with 1 by $\iota_{\mathrm{dR},p}$, it follows that

$$\iota_{\mathrm{dR},p}(c_{\mathrm{dR}}) = c_{\mathrm{crys}},$$

and (5.4.c) shows[8] that the induced action of ϕ_p on c_{dR} is just

$$\phi_p c_{\mathrm{dR}} = p c_{\mathrm{dR}}. \tag{5.4.d}$$

This implies that $(1 - p^{-1}\phi_p) c_{\mathrm{dR},H} \in H^1_{\mathrm{dR}}((Y, W)/\mathbb{Q}_p)$ is the image of something in $H^0_{\mathrm{dR}}(W/\mathbb{Q}_p)$. As $c_{\mathrm{dR},H} \in F^1 H^1_{\mathrm{dR}}((Y, W)/\mathbb{Q}_p)$ and $c_{\mathrm{dR}} \in F^1 H^1_{\mathrm{dR}}(Y/\mathbb{Q}_p)$ and since $\alpha_{\mathrm{dR}} : F^1 H^1_{\mathrm{dR}}((Y, W)/\mathbb{Q}_p) \xrightarrow{\sim} F^1 H^1_{\mathrm{dR}}(Y/\mathbb{Q}_p)$, this 'something' is nonzero modulo the image of $H^0_{\mathrm{dR}}(Y/\mathbb{Q}_p)$ if and only if the bottom of (5.4.a) is a nonsplit extension of filtered $\mathbb{Q}_p$-spaces equipped with a $\mathbb{Q}_p[\phi_p]$-module structure.

[8] This also follows as the spaces being compared are one-dimensional, but this argument works in more general settings.

Ideally there would be $\omega' \in H^0(\Omega^1_{X/\mathbb{Q}_p}(\log Z))$ such that $(1-p^{-1}\phi_p)c_{\mathrm{dR},H} = [\omega']_W$. As $0 = [\omega'] \in H^1_{\mathrm{dR}}(Y/\mathbb{Q}_p)$, it would have to be that $\omega' = d\eta$ for some $\eta \in H^0(Y, \mathcal{O}_{Y/\mathbb{Q}_p})$ and hence that $(1-p^{-1}\phi_p)c_{\mathrm{dR},H}$ is the image of $\eta|_W$. The nonvanishing of this image (and so of the class $\mathcal{E}_{F,\text{ét}}|_{G_{\mathbb{Q}_p}}$) would be equivalent to $\lambda(\eta|_W) \neq 0$ for some $\mathbb{Q}_p$-homomorphism $\lambda : H^0_{\mathrm{dR}}(W/\overline{\mathbb{Q}}_p) \to \overline{\mathbb{Q}}_p$ that vanishes on the image of $H^0_{\mathrm{dR}}(Y/\mathbb{Q}_p)$. Unfortunately, this ideal situation does not hold in general. However, we can essentially realize it by passing from algebraic de Rham cohomology to another cohomology theory, one where the whole of the cohomology group $H^1_{\mathrm{dR}}((Y, W)/\mathbb{Q}_p)$ can be represented by differentials, much as $H^1_{\mathrm{dR}}((Y, W)/\mathbb{C})$ can be represented by real analytic differentials.

5.5. ***Monsky–Washnitzer cohomology.*** Let W, X, Y, etc., be the special fibers of $\mathscr{W}$, $\mathscr{X}$, $\mathscr{Y}$, etc. The de Rham cohomology groups on the top line of (5.4.a) are naturally identified with the corresponding Monsky–Washnitzer (MW) cohomology of the corresponding special fibers, yielding a commutative diagram

$$\begin{array}{ccccc} \dfrac{H^0_{\mathrm{dR}}(W/\mathbb{Q}_p)}{\mathrm{im}(H^0_{\mathrm{dR}}(Y/\mathbb{Q}_p))} & \hookrightarrow & H^1_{\mathrm{dR}}((Y, W)/\mathbb{Q}_p) & \xrightarrow{\alpha_{\mathrm{dR}}} & H^1_{\mathrm{dR}}(Y/\mathbb{Q}_p) \\ \Vert & & \Vert & & \Vert \\ \dfrac{H^0_{\mathrm{MW}}(\mathsf{W},\mathbb{Q}_p)}{\mathrm{im}(H^0_{\mathrm{MW}}(\mathsf{Y},\mathbb{Q}_p))} & \hookrightarrow & H^1_{\mathrm{MW}}(\mathsf{Y}, \mathsf{W}, \mathbb{Q}_p) & \xrightarrow{\alpha_{\mathrm{MW}}} & H^1_{\mathrm{MW}}(\mathsf{Y}, \mathbb{Q}_p). \end{array} \tag{5.5.a}$$

The MW cohomology groups are defined as follows. Let

$$A_0^\dagger = \mathbb{Z}_p\langle t, x\rangle^\dagger/((t-1)x-1)$$

be the weak completion of $A_0 = \mathbb{Z}_p\big[t, \frac{1}{t-1}\big]$ and let

$$\Omega^1_{A_0^\dagger} = (A_0^\dagger dt + A_0^\dagger dx)/A_0^\dagger(x\,dt + (t-1)dx)$$

be the module of continuous differentials. Here $\mathbb{Z}_p\langle t, x\rangle^\dagger$ consists of the power series $\sum_{n,m=0}^\infty a_{n,m}t^n x^m$, $a_{n,m} \in \mathbb{Z}_p$, for which there exists a constant $C > 0$ and a real number $0 < \rho < 1$ such that $|a_{n,m}|_p \le C\rho^{n+m}$ for all n, m. Let $A^\dagger = A_0^\dagger \otimes_{\mathbb{Z}_p} \mathbb{Q}_p$ and

$$\Omega^1_{A^\dagger} = \Omega^1_{A_0^\dagger} \otimes_{\mathbb{Z}_p} \mathbb{Q}_p.$$

Then the cohomology group $H^1_{\mathrm{MW}}(\mathsf{Y}, \mathbb{Q}_p)$ is canonically computed by the cohomology of the complex $DR_Y^\dagger = [A^\dagger \xrightarrow{d} \Omega^1_{A^\dagger}]$. Similarly, $H^1_{\mathrm{MW}}(\mathsf{Y}, \mathsf{W}, \mathbb{Q}_p)$ is computed by the cohomology of the complex $DR^\dagger_{\mathsf{Y}}(-\mathsf{W}) = [\Phi_N(t)A^\dagger \to \Omega^1_{A^\dagger}]$, and so

$$H^1_{\mathrm{MW}}(\mathsf{Y}, \mathbb{Q}_p) = \Omega^1_{A^\dagger}/dA^\dagger \quad \text{and} \quad H^1_{\mathrm{MW}}(\mathsf{Y}, \mathsf{W}, \mathbb{Q}_p) = \Omega^1_{A^\dagger}/d(\Phi_N(t)A^\dagger).$$

The maps between the top and bottom rows of (5.5.a) are induced by the obvious maps of complexes $DR_Y \to DR^\dagger_{\mathsf{Y}}$ and $DR_Y(-W) \to DR^\dagger_{\mathsf{Y}}(-\mathsf{W})$. In particular, the map from $F^1H^1_{\mathrm{dR}}(Y/\mathbb{Q}_p) = \Omega^1_A/dA$ to $H^1_{\mathrm{MW}}(\mathrm{Y}, \mathbb{Q}_p)$ is just the obvious one, and similarly for $F^1H^1_{\mathrm{dR}}((Y, W)/\mathbb{Q}_p) = \Omega^1_A/d(\Phi_N(t)A)$.

The Monsky–Washnitzer cohomology groups are also equipped with a canonical Frobenius action induced by any homomorphism $F_p : A_0^\dagger \to A_0^\dagger$ that reduces mod p to the usual Frobenius map on $A_0/pA_0 = A_0^\dagger/pA_0^\dagger$. In this case there is a unique such F_p that sends t to t^p. The canonical Frobenius action on the Monsky–Washnitzer cohomology groups agrees with the Frobenius action ϕ_p on the de Rham cohomology groups. For more on Monsky–Washnitzer cohomology and the various objects introduced above, the interested reader should see [van der Put 1986].

Our problem now becomes one of finding an explicit $\eta \in A^\dagger$ such that $d\eta = (1 - p^{-1}F_p^*)\omega$. To do this we enlarge the class of functions we are working with.

5.6. ***Coleman integration and the calculation.*** The ring $A^\dagger$ is a subring of the rigid analytic functions on the affinoid $Y_{\mathrm{an}} = \mathrm{spm}(\mathcal{A})$ for $\mathcal{A} = \mathbb{Q}_p\langle t, x\rangle/((t-1)x-1))$, where $\mathbb{Q}_p\langle t, x\rangle$ is the standard Tate algebra. The geometric points of Y_{an} comprise the set

$$Y_{\mathrm{an}}(\overline{\mathbb{Q}}_p) = \{t \in \mathcal{O}_{\overline{\mathbb{Q}}_p} : |t-1|_p = 1\},$$

for $\mathcal{O}_{\overline{\mathbb{Q}}_p}$ the ring of integers of $\overline{\mathbb{Q}}_p$. That is, $Y_{\mathrm{an}}(\overline{\mathbb{Q}}_p)$ is $\mathcal{O}_{\overline{\mathbb{Q}}_p}$ with the open disc of radius 1 around 1 removed. The above identification just sends a $\mathbb{Q}_p$-homomorphism $\mathcal{A} \twoheadrightarrow \mathcal{A}/\mathfrak{m} \hookrightarrow \overline{\mathbb{Q}}_p$, $\mathfrak{m} \in \mathrm{spm}(\mathcal{A})$, to the image of t under this homomorphism. The ring $A^\dagger$ is then identified with a subring of the locally analytic functions $\mathcal{A}_{\mathrm{loc}}$ on Y_{an} over $\overline{\mathbb{Q}}_p$, where $\mathcal{A}_{\mathrm{loc}}$ is the ring of $\overline{\mathbb{Q}}_p$-valued functions $f(t)$ on the set $\{t \in \mathcal{O}_{\overline{\mathbb{Q}}_p} : |t-1|_p = 1\}$ such that (i) $\sigma(f(t)) = f(\sigma(t))$ for all $\sigma \in \mathrm{Gal}(\overline{\mathbb{Q}}_p/L)$ for some finite extension $L/\mathbb{Q}_p$ and (ii) on some open disc $\{t \in \mathcal{O}_{\overline{\mathbb{Q}}_p} : |t-t_0|_p < \epsilon\}$ around each point t_0, $f(t)$ is equal to a convergent power series in $t - t_0$. There is an obvious notion of locally analytic differentials on $Y_{\mathrm{an}}(\overline{\mathbb{Q}}_p)$ over $\overline{\mathbb{Q}}_p$ and we denote the $\mathcal{A}_{\mathrm{loc}}$-module of such by $\Omega^1_{\mathcal{A}_{\mathrm{loc}}}$. There is also an induced embedding $\Omega^1_{A^\dagger} \otimes_{\mathbb{Q}_p} \overline{\mathbb{Q}}_p \hookrightarrow \Omega^1_{\mathcal{A}_{\mathrm{loc}}}$, which is compatible with the differentials

$$d : A^\dagger \otimes_{\mathbb{Q}_p} \overline{\mathbb{Q}}_p \to \Omega^1_{A^\dagger} \quad \text{and} \quad d : \mathcal{A}_{\mathrm{loc}} \to \Omega^1_{\mathcal{A}_{\mathrm{loc}}}.$$

We will make use of Coleman integration (see [Besser 2012]), which is a $\overline{\mathbb{Q}}_p$-linear map $\int : \Omega^1_{A^\dagger} \otimes_{\mathbb{Q}_p} \overline{\mathbb{Q}}_p \to \mathcal{A}_{\mathrm{loc}}/\overline{\mathbb{Q}}_p$, to determine η:

$$\eta = \int (1 - p^{-1}F_p^*)\omega \in A_{\mathrm{loc}}/\overline{\mathbb{Q}}_p.$$

Note that η is only well-defined up to the addition of a constant, an ambiguity that does not affect the value $\lambda(\eta|_W)$.

The Frobenius F_p on $\mathcal{A}^\dagger$ is the restriction of $F_{p,\mathrm{loc}} : Y_{\mathrm{an}}(\overline{\mathbb{Q}}_p) \to Y_{\mathrm{an}}(\overline{\mathbb{Q}}_p)$, $t \mapsto t^p$, in the sense that $(F_pf)(t) = f(t^p)$ for $f \in A^\dagger$. The theory of Coleman integration provides a unique $\overline{\mathbb{Q}}_p$-linear map $\int : \Omega^1_{A^\dagger} \otimes_{\mathbb{Q}_p} \overline{\mathbb{Q}}_p \to \mathcal{A}_{\mathrm{loc}}/\overline{\mathbb{Q}}_p$ such that (a) $d \circ \int : \Omega^1_{A^\dagger} \otimes_{\mathbb{Q}_p} \overline{\mathbb{Q}}_p \hookrightarrow \Omega^1_{\mathcal{A}_{\mathrm{loc}}}$ and $\int \circ d : A^\dagger \otimes_{\mathbb{Q}_p} \overline{\mathbb{Q}}_p \to \mathcal{A}_{\mathrm{loc}}/\overline{\mathbb{Q}}_p$ are the canonical maps, and (b) $F_p^* \circ \int = \int \circ F_p^*$. The condition (b) actually holds for *any* lift F_p of the Frobenius map. Not surprisingly, it is relatively straightforward to show that $\int \frac{dt}{t-1} = \log_p(1-t)$, where $\log_p$ is the usual Iwasawa branch of the p-adic logarithm (so $\log_p(p) = 0$); see [Besser 2012, §1.2]. It then follows from (b) that

$$\eta = -\log_p(1-t) + p^{-1} \log_p(1-t^p) \in \mathcal{A}_{\mathrm{loc}}/\overline{\mathbb{Q}}_p.$$

Let $\chi : (\mathbb{Z}/N\mathbb{Z})^\times \to \overline{\mathbb{Q}}_p^\times$ be any nontrivial character. Then

$$\lambda_{\chi,\mathrm{dR}} : H^0_{\mathrm{dR}}(W/\overline{\mathbb{Q}}_p) \to \mathbb{Q}_p, \quad \lambda_{\chi,\mathrm{dR}}((x_\zeta)_{\zeta \in W(\overline{\mathbb{Q}}_p)}) = \frac{1}{\tau(\chi_0)} \sum_{a \in (\mathbb{Z}/N\mathbb{Z})^\times} \chi(a) x_{\zeta_N^a},$$

is 0 on the image of $H^0_{\mathrm{dR}}(Y/\overline{\mathbb{Q}}_p)$, which is the image of the diagonal embedding $\overline{\mathbb{Q}}_p \hookrightarrow \bigoplus_{\zeta \in W(\overline{\mathbb{Q}}_p)} \overline{\mathbb{Q}}_p$. Here χ_0 is the primitive Dirichlet character associated to χ and $\tau(\chi_0) = \sum_{a \in (\mathbb{Z}/N_0\mathbb{Z})^\times} \chi_0(a) \zeta_{N_0}^a$ is its usual Gauss sum. Then

$$\begin{aligned}
\lambda_{\chi,\mathrm{dR}}(\eta|_W) &= -\frac{1}{\tau(\chi_0)} \sum_{a \in (\mathbb{Z}/N\mathbb{Z})^\times} \chi(a) \eta(\zeta_N^a) \\
&= -\frac{1}{\tau(\chi_0)} (1 - \bar{\chi}(p) p^{-1}) \sum_{a \in (\mathbb{Z}/N\mathbb{Z})^\times} \chi(a) \log_p(1 - \zeta_N^a).
\end{aligned}$$

If χ is odd (so $\chi(-a) = -\chi(a)$), then the last sum vanishes, as $\log_p(1 - \zeta_N^{-a}) = \log_p(-\zeta_N^{-a}(1 - \zeta_N^a)) = \log_p(1 - \zeta_N^a)$. But if χ is even (so $\chi(a) = \chi(-a)$), then the sum equals

$$L_p(1, \bar{\chi}_0) \prod_{\substack{\ell \text{ prime} \\ \ell \mid N \\ \ell \nmid N_0}} (1 - \chi_0(\ell))$$

by a well-known formula for the value of the p-adic Dirichlet L-function $L_p(s, \bar{\chi}_0)$ at the point $s = 1$ (see [Washington 1997, Theorem 5.18]). Here, as before, N_0 is the conductor of χ_0. As $L_p(1, \bar{\chi}_0) \neq 0$ (see [Washington 1997, Corollary 5.30]) we see — just as in the complex case — that $\lambda_{\chi,\mathrm{dR}}(\eta|_W)$ is nonzero if and only if $\chi_0(\ell) \neq 1$ for all $\ell \mid N$, $\ell \nmid N_0$. And, also as before, this is equivalent to $\mathrm{ord}_{s=0} L(s, \chi) = 1$. Hence the nonvanishing of $\lambda_{\chi,\mathrm{dR}}(\eta|_W)$ also agrees with $\mathrm{ord}_{s=0} L(s, \chi) = 1$.

As noted before, $\mathcal{E}_{\mathbb{Q}_p,\text{ét}}|_{G_{\mathbb{Q}_p}}$ is a nonsplit extension of p-adic Galois representations if and only if $\lambda(\eta|_W) \neq 0$ for some nonzero $\lambda : H^0_{\mathrm{dR}}(W/\mathbb{Q}_p) \to \overline{\mathbb{Q}}_p$ that vanishes on the image of $H^0_{\mathrm{dR}}(Y/\mathbb{Q}_p)$. Such λ are exactly the nonzero linear combinations of the $\lambda_{\mathrm{dR},\chi}$ for χ running over the nontrivial characters of $(\mathbb{Z}/N\mathbb{Z})^\times$.

So as a consequence of the calculation above we have:

$$\boxed{\begin{array}{l}\text{there exists a nontrivial even character}\\ \qquad\chi : (\mathbb{Z}/N\mathbb{Z})^\times \to \mathbb{C}^\times \\ \quad\text{such that } \mathrm{ord}_{s=0} L(s,\chi) = 1\end{array} \iff \mathcal{E}_{\mathbb{Q}_p,\text{ét}}|G_{\mathbb{Q}_p} \neq 0.} \tag{5.6.a}$$

The left-hand side is, of course, satisfied if there is a primitive even character modulo N.

Just as in the case of extensions of Hodge structures, this can be refined. Suppose χ is $\mathbb{Q}_p^\times$-valued (which holds, for example, if $\phi(N) \mid (p-1)$). Then

$$\lambda_{\chi,\text{ét}} : H^0_{\text{ét}}(\overline{W}, \mathbb{Q}_p) \to \mathbb{Q}_p(\chi), \quad \lambda_{\chi,\text{ét}}((x_\zeta)_{\zeta\in W(\overline{\mathbb{Q}})}) = \sum_{a\in(\mathbb{Z}/N\mathbb{Z})^\times} \chi(a)x_{\zeta_N^a},$$

is a $\mathbb{Q}_p[G_\mathbb{Q}]$-module homomorphism. Here we view $\mathbb{Q}_p(\chi)$ as $\mathbb{Q}_p$ but with $G_\mathbb{Q}$ action via the Galois character χ. So $1 \in \mathbb{Q}_p(\chi)$ is a $\mathbb{Q}_p$-basis and $\sigma\cdot 1 = \chi(\sigma)\cdot 1 = \chi(\sigma)$. It follows that $\mathcal{E}_{\chi,\text{ét}} = H^1_{\text{ét}}(\overline{Y}, \overline{W}, \mathbb{Q}_p)/\ker(\lambda_\chi)$ is an extension of $\mathbb{Q}_p[G_\mathbb{Q}]$-modules that fits into a commutative diagram:

$$\begin{tikzcd} \frac{H^0_{\text{ét}}(\overline{W},\mathbb{Q}_p)}{\mathrm{im}(H^0_{\text{ét}}(\overline{Y},\mathbb{Q}_p))} \arrow[r, hook] \arrow[d, two heads, "\lambda_{\chi,\text{ét}}"] & H^1_{\text{ét}}(\overline{Y}, \overline{W}, \mathbb{Q}_p) \arrow[r, two heads] \arrow[d, two heads, "/\ker(\lambda_{\chi,\text{ét}})"] & H^1_{\text{ét}}(\overline{Y}, \mathbb{Q}_p) \arrow[d, equal] \\ \mathbb{Q}_p(\chi) \arrow[r, hook] & \mathcal{E}_{\chi,\text{ét}} \arrow[r, two heads] & \mathbb{Q}_p c_{\text{ét}} = \mathbb{Q}_p(-1). \end{tikzcd} \tag{5.6.b}$$

In particular, $\mathcal{E}_{\chi,\text{ét}} \in \mathrm{Ext}^1_{\mathbb{Q}_p[G_\mathbb{Q}]}(\mathbb{Q}(\chi), \mathbb{Q}c_{\text{ét}}) = \mathrm{Ext}^1_{\mathbb{Q}_p[G_\mathbb{Q}]}(\mathbb{Q}_p(\chi), \mathbb{Q}_p(-1))$. The calculation above shows that

$$\boxed{\chi \text{ even and nontrivial, } \mathrm{ord}_{s=0} L(s,\chi) = 1 \iff \mathcal{E}_{\chi,\text{ét}}|G_{\mathbb{Q}_p} \neq 0.} \tag{5.6.c}$$

5.6.1. Remark. The fact that $\mathcal{E}_{\chi,\text{ét}}|G_{\mathbb{Q}_p} = 0$ if χ is odd is consistent with the fact that $L(0,\chi) \neq 0$ for χ odd and primitive, and so we do not expect extensions.

5.6.2. Remark. A careful reader may have noted that the definitions of $\lambda_{\chi,\mathrm{dR}}$ and $\lambda_{\chi,\text{ét}}$ differ by a factor of $\tau(\chi_0)$. This difference is partly explained by the commutativity of

$$\begin{tikzcd} H^0_{\mathrm{dR}}(W/\mathbb{Q}_p) \arrow[r, "\iota_{\mathrm{dR},p}"] \arrow[d, two heads, "\lambda_{\chi,\mathrm{dR}}"] & D_{\mathrm{dR}}(H^0_{\text{ét}}(\overline{W}, \mathbb{Q}_p)) \arrow[d, two heads, "\lambda_{\chi,\text{ét}}\otimes\mathrm{id}"] \\ \mathbb{Q}_p \arrow[r, "a\mapsto a(1\otimes\tau(\chi_0))"] & D_{\mathrm{dR}}(\mathbb{Q}_p(\chi)). \end{tikzcd}$$

Note that $D_{\mathrm{dR}}(\mathbb{Q}_p(\chi)) = \mathbb{Q}_p(1\otimes\tau(\chi_0)) \subset \mathbb{Q}_p(\chi)\otimes_{\mathbb{Q}_p} B_{\mathrm{dR}}$. This relation figures into the derivation of the expression for the Bloch–Kato logarithm given in the supplement below.

5.6.3. Remark. The careful reader may also have noted that we have not fully succeeded in avoiding special units: the formula for $L_p(1, \bar{\chi}_0)$ involves p-adic logs of what are essentially cyclotomic units (and similarly for $L(1, \bar{\chi}_0)$ in the Hodge case). But even this can be avoided by working with modular curves in place of the projective line, as explained in the example from Section 6.1 below.

5.7. *Vista: the Bloch–Kato logarithm.* The extension $\mathcal{E}_{\chi,\text{ét}}$ determines a class $z_\chi \in H^1(\mathbb{Q}, \mathbb{Q}_p(\chi\epsilon))$ as follows: Take the 1-Tate twist of the extension $\mathbb{Q}_p(\chi) \hookrightarrow \mathcal{E}_{\chi,\text{ét}} \twoheadrightarrow \mathbb{Q}_p(-1)\ (= \mathbb{Q}_p c_{\text{ét}})$. This gives an extension

$$\mathbb{Q}_p(\chi\epsilon) \hookrightarrow \mathcal{E}_{\chi,\text{ét}}(1) \twoheadrightarrow \mathbb{Q}_p\ (= \mathbb{Q}_p(c_{\text{ét}} \otimes \underline{\zeta})).$$

Here we have identified $\mathbb{Q}_p(\chi)(1)$ with $\mathbb{Q}_p(\chi\epsilon)$ using the basis $1 \otimes \underline{\zeta} \in \mathbb{Q}_p(\chi)(1)$. Let $\tilde{c} \in \mathcal{E}_{\chi,\text{ét}}(1)$ be any element mapping to $c_{\text{ét}} \otimes \underline{\zeta}$. Then z_χ is just the class of the 1-cocycle $\sigma \mapsto \sigma\tilde{c} - \tilde{c}$. The class z_χ is just the image of $c_{\text{ét}} \otimes \underline{\zeta}$ under the boundary map $\mathbb{Q}_p(c_{\text{ét}} \otimes \underline{\zeta}) \to H^1(\mathbb{Q}, \mathbb{Q}_p(\chi\epsilon))$ of the long exact cohomology sequence associated with the short exact sequence displayed above.

Assuming (5.4.b), we showed that the restriction of $\mathcal{E}_{\chi,\text{ét}}$ to $G_{\mathbb{Q}_p}$ is nontrivial, provided some value of a p-adic L-function is nonzero. This nontriviality is equivalent to $\mathrm{loc}_p(z_\chi) \in H^1(\mathbb{Q}_p, \mathbb{Q}_p(\chi\epsilon))$ being nonzero. As the extension $\mathcal{E}_\chi|_{G_{\mathbb{Q}_p}}$ is a crystalline extension, so is its 1-Tate twist. Hence $\mathrm{loc}_p(z_\chi)$ belongs to the Bloch–Kato subspace

$$H^1_f(\mathbb{Q}_p, \mathbb{Q}_p(\chi\epsilon)) = \ker\bigl\{H^1(\mathbb{Q}_p, \mathbb{Q}_p(\chi\epsilon)) \to H^1(\mathbb{Q}_p, \mathbb{Q}_p(\chi\epsilon) \otimes_{\mathbb{Q}_p} B_{\text{crys}})\bigr\}.$$

This group is computed by the extended Bloch–Kato exponential

$$\widetilde{\exp}_{\text{BK}} : \frac{D_{\text{crys}}(\mathbb{Q}_p(\epsilon\chi)) \oplus \bigl(D_{\text{dR}}(\mathbb{Q}_p(\chi\epsilon))/F^0 D_{\text{dR}}(\mathbb{Q}_p(\chi\epsilon))\bigr)}{\bigl\{\bigl((1-\phi_p)x, x \bmod F^0 D_{\text{dR}}(\mathbb{Q}_p(\chi\epsilon))\bigr) : x \in D_{\text{crys}}(\mathbb{Q}_p(\chi\epsilon))\bigr\}} \xrightarrow{\sim} H^1_f(\mathbb{Q}_p, \mathbb{Q}_p(\chi\epsilon)),$$

which is a boundary map in the long-exact sequence of $G_{\mathbb{Q}_p}$-cohomology for the tensor product over $\mathbb{Q}_p$ of $\mathbb{Q}_p(\chi\epsilon)$ with the short exact sequence $\mathbb{Q}_p \hookrightarrow B_{\text{crys}} \twoheadrightarrow B_{\text{crys}} \oplus (B_{\text{dR}}/B^+_{\text{dR}})$, the last arrow being $x \mapsto ((1-\phi_p x), x \bmod B^+_{\text{dR}})$. The inverse of this is the Bloch–Kato logarithm. As $\mathbb{Q}_p(\chi\epsilon)$ is a crystalline representation of $G_{\mathbb{Q}_p}$ (assuming (5.4.b)), $D_{\text{crys}}(\mathbb{Q}_p(\chi\epsilon)) = D_{\text{dR}}(\mathbb{Q}_p(\chi\epsilon))$, so the restriction of $\widetilde{\exp}_{\text{BK}}$ to the $D_{\text{crys}}(\mathbb{Q}_p(\epsilon\chi))$ summand induces an isomorphism

$$\widetilde{\exp}_{\text{BK}} : \frac{D_{\text{crys}}(\mathbb{Q}_p(\epsilon\chi))}{\bigl\{(1-\phi_p)x : x \in F^0 D_{\text{crys}}(\mathbb{Q}_p(\chi\epsilon))\bigr\}} \xrightarrow{\sim} H^1_f(\mathbb{Q}_p, \mathbb{Q}_p(\chi\epsilon)).$$

In this particular case, $F^0 D_{\text{crys}}(\mathbb{Q}_p(\chi\epsilon)) = F^0 D_{\text{dR}}(\mathbb{Q}_p(\chi\epsilon)) = 0$, so we have

$$\widetilde{\exp}_{\text{BK}} : D_{\text{crys}}(\mathbb{Q}_p(\chi\epsilon)) \xrightarrow{\sim} H^1_f(\mathbb{Q}_p, \mathbb{Q}_p(\chi\epsilon)).$$

Let $\widetilde{\log}_{\mathrm{BK}} : H^1_f(\mathbb{Q}_p, \mathbb{Q}_p(\chi\epsilon)) \xrightarrow{\sim} D_{\mathrm{crys}}(\mathbb{Q}_p(\chi\epsilon))$ be the inverse of $\widetilde{\exp}_{\mathrm{BK}}$. It is natural to ask whether we can identify the element $\lambda_{\mathrm{crys}} \in D_{\mathrm{crys}}(\mathbb{Q}_p(\epsilon\chi))$ such that $\widetilde{\log}_{\mathrm{BK}}(\mathrm{loc}_p(z_\chi)) = \lambda_{\mathrm{crys}}$. As it turns out, we have already computed this:

$$D_{\mathrm{crys}}(\mathbb{Q}_p(\chi\epsilon)) = D_{\mathrm{dR}}(\mathbb{Q}_p(\chi\epsilon)) = D_{\mathrm{dR}}(\mathbb{Q}_p(\chi)(1)) = \mathbb{Q}_p(1 \otimes \underline{\zeta} \otimes \underline{t}^{-1})$$

and

$$\boxed{\widetilde{\log}_{\mathrm{BK}}(\mathrm{loc}_p(z_\chi)) = L_p(1, \bar{\chi}_0) \prod_{\substack{\ell \text{ prime} \\ \ell \mid N \\ \ell \nmid N_0}} (1 - \chi_0(\ell)) \cdot (1 \otimes \underline{\zeta} \otimes \underline{t}^{-1}).} \tag{5.7.a}$$

So the Bloch–Kato logarithm of $\mathrm{loc}_p(z_\chi)$ is naturally identified with the value of a p-adic L-function.

The equality in (5.7.a) can be seen as follows. For crystalline $\mathbb{Q}_p$-representations V of $G_{\mathbb{Q}_p}$, the groups $H^0(\mathbb{Q}_p, V)$ and $H^1_f(\mathbb{Q}_p, V)$ are functorially computed by the complex $C_{\mathrm{crys}}(V) = [D_{\mathrm{crys}}(V) \to D_{\mathrm{crys}}(V) \oplus D_{\mathrm{dR}}(V)/F^0 D_{\mathrm{dR}}(V)]$, where the arrow is the map $x \mapsto ((1 - \phi_p)x, x \bmod F^0 D_{\mathrm{dR}}(V))$. Applying this to the two sequences in the 1-Tate twist of the commutative diagram (5.6.b), employing the snake lemma to compute the boundary map

$$\begin{aligned} H^0(C_{\mathrm{crys}}(\mathbb{Q}_p)) &= H^0(C_{\mathrm{crys}}(\mathbb{Q}_p(c_{\text{ét}} \otimes \underline{\zeta}))) \\ &\to H^1(C_{\mathrm{crys}}(\lambda_{\chi,\text{ét}}(H^0_{\text{ét}}(\overline{W})(1))) = H^1(C_{\mathrm{crys}}(\mathbb{Q}_p(\chi\epsilon))), \end{aligned}$$

and appealing to the relation in Remark 5.6.2 yields the displayed formula for $\widetilde{\log}_{\mathrm{BK}}(\mathrm{loc}_p(z_\chi))$, which is just the image of $c_{\text{ét}} \otimes \underline{\zeta} \otimes \underline{t}^{-1} \in H^0(C_{\mathrm{crys}}(\mathbb{Q}_p(c_{\text{ét}} \otimes \underline{\zeta})))$ under the above boundary map.

5.8. *Vista: Euler systems.* A variation on the definition of the classes z_χ yields an Euler system. For a reader with some familiarity with Euler systems this should not be surprising in light of the relation (5.7.a). Recall that we are assuming that χ is nontrivial and $\mathbb{Q}_p$-valued and that $p \nmid N$ (all for simplicity).

First we note that we can replace $H^1_{\text{ét}}(\overline{Y}, \mathbb{Q}_p)$ with $H^1_{\text{ét}}(\overline{Y}, \mathbb{Z}_p)$ in the definition of $c_{\text{ét}}$. So in particular, $z_\chi \in H^1(\mathbb{Q}, \mathbb{Z}_p(\chi\epsilon))$ with $\mathbb{Z}_p(\chi\epsilon)$ the free $\mathbb{Z}_p$-module of rank one with $\sigma \in G_{\mathbb{Q}}$ acting via multiplication by $\chi\epsilon(\sigma)$. The other classes of our Euler system come from slightly modifying the definition of $\mathcal{E}_{\chi,\text{ét}}$. For each integer M such that $(N, M) = 1$ we let $Z_M = \mu_M \cup \{\infty\} \subset X$ and $Y_M = X \setminus Z_M$. Note that $W \subset Y_M$. Note also that we recover Y by taking $M = 1$. Then just as in Section 5.3 we have $H^1_{\text{ét}}(\overline{Y}_M, \mathbb{Z}_p) \hookrightarrow H^0(\overline{Z}_M, \mathbb{Z}_p(-1)) = \bigoplus_{z \in Z_M(\overline{\mathbb{Q}})} \mathbb{Z}_p(-1)$ with image equal to $\big\{(a_z \otimes \underline{\zeta}^\vee)_{z \in Z_M(\overline{\mathbb{Q}})} : \sum_z a_z = 0\big\}$. For $\zeta \in \mu_M$ we let $c_{\text{ét},\zeta} \in H^1_{\text{ét}}(\overline{Y}_M, \mathbb{Z}_p)$ be the class corresponding to $a_\infty = 1$, $a_\zeta = -1$, and $a_z = 0$ otherwise. The Galois group $G_{\mathbb{Q}}$ acts on $c_{\text{ét},\zeta}$ as $\sigma c_{\text{ét},\zeta} = \epsilon(\sigma)^{-1} c_{\text{ét},\sigma(\zeta)}$. In particular, $G_{\mathbb{Q}[\mu_M]}$

acts on $c_{\text{ét},\zeta}$ as multiplication by ϵ^{-1}. That is, $\mathbb{Z}_p c_{\text{ét},\zeta} \simeq \mathbb{Z}_p(-1)$ as a $\mathbb{Z}_p[G_{\mathbb{Q}[\mu_M]}]$-module. Pulling back to $H^1_{\text{ét}}(\overline{Y}_M, \overline{W}, \mathbb{Z}_p)$ and then pushing out by λ_χ as before yields an extension $\mathcal{E}_{\chi,\zeta} \in \operatorname{Ext}^1_{\mathbb{Z}_p[G_{\mathbb{Q}[\mu_M]}]}(\mathbb{Z}_p(-1), \mathbb{Z}_p(\chi))$ and hence a class $z_{\chi,\zeta} \in H^1(\mathbb{Q}[\mu_M], \mathbb{Z}_p(\chi\epsilon))$. Note that if $\zeta \in \mu_{M'}$ for some $M' \mid M$, then these are just the restrictions to $G_{\mathbb{Q}[\mu_M]}$ of the extension and class defined with M' in place of M. Furthermore, it follows from the action of $G_{\mathbb{Q}}$ on the $c_{\text{ét},\chi}$ and the action of $G_{\mathbb{Q}}$ on $H^1(\mathbb{Q}[\mu_M], \mathbb{Z}_p(\chi\epsilon))$ (particularly in terms of cycle representatives) that

$$\sigma z_{\chi,\zeta} = z_{\chi,\sigma(\zeta)}. \tag{5.8.a}$$

As before, let $\zeta_M = e^{2\pi i/M} \in \mu_M$. We now set

$$z_{\chi,M} = \bar{\chi}(M) z_{\chi,\zeta_M} \in H^1(\mathbb{Q}[\mu_M], \mathbb{Z}_p(\chi\epsilon)).$$

It should not be surprising that

$$\boxed{\left\{z_{\chi,M} \in H^1(\mathbb{Q}[\mu_M], \mathbb{Z}_p(\chi\epsilon)) : (M, N) = 1\right\} \text{ is an Euler system.}}$$

Here, by an Euler system we mean a collection of cohomology classes as in [Rubin 2000]. In particular, the $z_{\chi,M}$ satisfy the norm relations

$$\boxed{\operatorname{cor}_{\mathbb{Q}[\mu_{M\ell}]/\mathbb{Q}[\mu_M]} z_{\chi,M\ell} = \begin{cases} (1 - \bar{\chi}(\ell)\operatorname{frob}_\ell^{-1}) z_{\chi,M}, & \ell \nmid NMp, \\ z_{\chi,M}, & \ell \mid M. \end{cases}} \tag{5.8.b}$$

We quickly explain how to see these relations.

Since the restriction map $H^1(\mathbb{Q}[\mu_M], \mathbb{Z}_p(\chi\epsilon)) \hookrightarrow H^1(\mathbb{Q}[\mu_{M\ell}], \mathbb{Z}_p(\chi\epsilon))$ is an injection, it is enough to check that the equality of the norm relation holds in $H^1(\mathbb{Q}[\mu_{M\ell}], \mathbb{Z}_p(\chi\epsilon))$. From (5.8.a) we see that

$$\begin{aligned} &\operatorname{cor}_{\mathbb{Q}[\mu_{M\ell}]/\mathbb{Q}[\mu_M]} z_{\chi,M\ell} \\ &\quad = \bar{\chi}(M\ell) \sum_{\sigma \in \operatorname{Gal}(\mathbb{Q}[\mu_{M\ell}]/\mathbb{Q}[\mu_M])} z_{\chi,\sigma(\zeta_{M\ell})} \in H^1(\mathbb{Q}[\mu_{M\ell}], \mathbb{Z}_p(\chi\epsilon)). \end{aligned} \tag{5.8.c}$$

We consider the map $f : Y_{M\ell} \to Y_M$, $f(t) = t^\ell$. This induces a commutative diagram

$$\begin{tikzcd}
 & & \mathbb{Z}_p c_{\chi,\zeta_M} \arrow[d, hook] \\
\dfrac{H^0_{\text{ét}}(\overline{W},\mathbb{Z}_p)}{\operatorname{im}(H^0_{\text{ét}}(\overline{Y}_M,\mathbb{Z}_p))} \arrow[r, hook] \arrow[d, two heads, "f^*"] & H^1_{\text{ét}}(\overline{Y}_M, \overline{W}, \mathbb{Z}_p) \arrow[r, two heads] \arrow[d, two heads, "f^*"] & H^1_{\text{ét}}(\overline{Y}_M, \mathbb{Z}_p) \arrow[d, "f^*"] \\
\dfrac{H^0_{\text{ét}}(\overline{W},\mathbb{Z}_p)}{\operatorname{im}(H^0_{\text{ét}}(\overline{Y}_{M\ell},\mathbb{Z}_p))} \arrow[r, hook] \arrow[d, two heads, "\lambda_{\chi,\text{ét}}"] & H^1_{\text{ét}}(\overline{Y}_{M\ell}, \overline{W}, \mathbb{Z}_p) \arrow[r, two heads] & H^1_{\text{ét}}(\overline{Y}_{M\ell}, \mathbb{Z}_p) \\
\mathbb{Z}_p(\chi) & &
\end{tikzcd}$$

It follows that the extension obtained by pulling back c_{χ,ζ_M} and pushing out by $\lambda_\chi \circ f^*$ is the same as that obtained by pulling back f^*c_{χ,ζ_M} and pushing out by λ_χ. As we have

$$\lambda_\chi \circ f^* = \chi(\ell)\lambda_\chi \quad \text{and} \quad f^*c_{\chi,\zeta_M} = \sum_{\zeta^\ell=\zeta_M} c_{\chi,\zeta},$$

it follows that

$$\chi(\ell)z_{\chi,M} = \sum_{\zeta^\ell=\zeta_M} z_{\chi,\zeta} = \begin{cases} \sum_{\sigma\in\mathrm{Gal}(\mathbb{Q}[\mu_{M\ell}]/\mathbb{Q}[\mu_M])} z_{\chi,\sigma(\zeta_{M\ell})} + z_{\chi,\zeta_M^{\bar{\ell}}}, & \ell\nmid M, \\ \sum_{\sigma\in\mathrm{Gal}(\mathbb{Q}[\mu_{M\ell}]/\mathbb{Q}[\mu_M])} z_{\chi,\sigma(\zeta_{M\ell})}, & \ell \mid M. \end{cases} \tag{5.8.d}$$

Here we have used that $\zeta_{M\ell} = \zeta_M^{\bar{\ell}}\zeta_\ell^{\bar{M}}$, where $\bar{\ell}\ell \equiv 1 \bmod M$ and $\bar{M}M \equiv 1 \bmod \ell$, and so $\sigma(\zeta_{M\ell}) = \zeta_M^{\bar{\ell}}\sigma(\zeta_\ell^{\bar{M}})$. Comparing (5.8.d) with (5.8.c) yields

$$\mathrm{cor}_{\mathbb{Q}[\mu_{M\ell}]/\mathbb{Q}[\mu_M]} z_{\chi,M\ell} = \begin{cases} \bar{\chi}(M)z_{\chi,\zeta_M} - \bar{\chi}(M\ell)z_{\chi,\zeta_M^{\bar{\ell}}}, & \ell\nmid M, \\ \bar{\chi}(M)z_{\chi,\zeta_M}, & \ell \mid M. \end{cases}$$

If $\ell\nmid MNp$, then z_{χ,ζ_M} is unramified at ℓ and $\mathrm{frob}_\ell^{-1} z_{\chi,\zeta_M} = z_{\chi,\mathrm{frob}_\ell^{-1}(\zeta_M)} = z_{\chi,\zeta_M^{\bar{\ell}}}$. The norm relations (5.8.b) follow.

5.8.1. Remark. There is nothing in this section that requires χ to be $\mathbb{Q}_p$-valued or N to be prime to p. One can replace $\mathbb{Z}_p$ with the ring of integers $\mathcal{O}$ for any finite extension of $\mathbb{Q}_p$ and take χ to be any nontrivial $\mathcal{O}$-valued Dirichlet character. The arguments carry over immediately. The trivial character can also be handled, albeit with some additional modification (to ensure that the chosen functional λ is still trivial on the image of $H^0(Y_M, \mathcal{O})$).

5.8.2. Remark. The proof of the norm relations we have given here — which may seem much more involved than that for cyclotomic units (see [Rubin 2000, III.2]) — provides a template for an approach that carries over to many other settings, such as in [Shang et al. ≥ 2024] and [Sangiovanni-Vincentelli and Skinner ≥ 2024a].

5.8.3. Remark. To obtain special value formulas from this (or any) Euler system one also needs to relate the restrictions to $G_{\mathbb{Q}_p}$ of the Euler system classes to values of a p-adic L-function, that is, prove a so-called explicit reciprocity law. This is essentially the point of the calculation in Section 5.7. The general case can be handled similarly. The only real obstacle to overcome is that if $p \mid M$ (or N) then the naive integral models $\mathscr{Y}_M$ and $\mathscr{X}$ of Y_M and X are not such that $\mathscr{Y}_M$ is the complement of a smooth (or even normal crossings) divisor in $\mathscr{X}$. But it is not hard to establish the existence of such models over $\mathbb{Z}_p[\mu_{p^r}]$ for $p^r \mid\mid M$. With this in hand, the arguments presented previously carry over with only slight modification.

6. Some variations, very briefly

The constructions in Section 5 can be viewed as a very special case of a general set-up. Indications of this are provided by the variations on the construction and analysis of $\mathcal{E}_{\chi,\text{ét}}$ described briefly in this section. These additional special cases can be used to recover the Euler system for Dirichlet characters and for Hecke characters of imaginary quadratic fields along with their connection with p-adic L-functions (see [Shang et al. $\geq$ 2024]). Though we do not include a discussion here, a simple variation on these constructions involving products of modular curves can be used to recover Kato's Euler system for an eigenform. Examples of new Euler systems (also with connections to p-adic L-functions) obtained using the same template are given in [Sangiovanni-Vincentelli and Skinner $\geq$ 2024a; $\geq$ 2024b].

6.1. ***Dirichlet characters (again).*** Let $N \geq 4$. Let $Y_1(N) \subset X_1(N)$ be the usual modular curves for the congruence subgroup $\Gamma_1(N)$, and let $C_1(N) = X \setminus Y$ be the cusps. These have models as smooth varieties over $\mathbb{Q}$. The cusps $C_1(N) = \Gamma_1(N) \setminus \mathbb{P}^1(\mathbb{Q})$ of $X_1(N) = \Gamma_1(N) \setminus [\mathbb{H} \sqcup \mathbb{P}^1(\mathbb{Q})]$ are in bijection with the set $\{(\bar{a}, \bar{c}) \in \mathbb{Z}/N\mathbb{Z} \times \mathbb{Z}/N\mathbb{Z} : (a, c, N) = 1\}/\sim$ where $(\bar{a}_1, \bar{c}_1) \sim (\bar{a}_2, \bar{c}_2) \Longleftrightarrow (\bar{a}_2, \bar{c}_2) = \pm(\bar{a}_1 + m\bar{c}_1, \bar{c}_1)$ for some $m \in \mathbb{Z}$. The bijection is given by $\mathbb{P}^1(\mathbb{Q}) \ni \left[\begin{smallmatrix} a \\ c \end{smallmatrix}\right] \mapsto (\bar{a}, \bar{c})$, $a, c \in \mathbb{Z}$, $(a, c) = 1$. When we write $\left[\begin{smallmatrix} a \\ c \end{smallmatrix}\right]$ for some element in $\mathbb{P}^1(\mathbb{Q})$ or the cusp it represents, we will always mean $a, c \in \mathbb{Z}$ and $(a, c) = 1$. Let $C_0 \subset C_1(N)$ be the set of cusps represented by some $\left[\begin{smallmatrix} a \\ c \end{smallmatrix}\right]$ with $(c, N) = 1$; there are $\frac{\phi(N)}{2}$ of them. Similarly, let $C_\infty \subset C_1(N)$ be the set of cusps represented by some $\left[\begin{smallmatrix} a \\ c \end{smallmatrix}\right]$ with $N \mid c$; there are also $\frac{\phi(N)}{2}$ of them. We take the models of $X_1(N)$ and $Y_1(N)$ over $\mathbb{Q}$ such that each cusp in C_∞ is defined over $\mathbb{Q}$ and each cusp $\left[\begin{smallmatrix} a \\ c \end{smallmatrix}\right]$ in C_0 is defined over $\mathbb{Q}[\mu_N]^+$: The action of $G_\mathbb{Q}$ on the cusps is such that if $\sigma \in G_\mathbb{Q}$ maps to $m \in (\mathbb{Z}/N\mathbb{Z})^\times = \mathrm{Gal}(\mathbb{Q}[\mu_N]/\mathbb{Q})$, then $\sigma \cdot \left[\begin{smallmatrix} a \\ c \end{smallmatrix}\right]$ is represented by $\left[\begin{smallmatrix} a \\ c' \end{smallmatrix}\right]$ with $c \equiv mc' \bmod N$. Note that C_0 and C_∞ are $\mathbb{Q}$-subvarieties of $X = X_1(N)$.

Let $\chi : (\mathbb{Z}/N\mathbb{Z})^\times \to \mathbb{C}^\times$ be a nontrivial, primitive, even Dirichlet character. There exists an Eisenstein series G_χ of weight 2 and level N with q-expansion

$$G_\chi(\tau) = \sum_{n=1}^{\infty} \biggl(\sum_{d \mid n} \bar{\chi}\Bigl(\frac{n}{d}\Bigr) d \biggr) q^n, \quad q = e^{2\pi i \tau}.$$

The constant term $c_P(G_\chi)$ of G_χ is 0 at any cusp $P \notin C_0$ and at $P = \left[\begin{smallmatrix} a \\ c \end{smallmatrix}\right] \in C_0$ it is $c_P(G_\chi) = \bar{\chi}(c)\tau(\bar{\chi})L(-1, \chi)/2N^2$. Let $\omega_\chi = G_\chi(\tau)d\tau$. This defines a holomorphic differential on $Y = X \setminus C_0$ with log poles along C_0. Let $\omega_\chi^{\text{an}} = \tau(\chi)\omega_\chi$. By considering the residues of the differential ω_χ^{an} at the cusps in C_0 (which are essentially the constant terms) and using that the Hecke eigenvalues of G_χ distinguish it from cuspforms, one can see that $c_\chi = \iota_{\text{dR}}([\omega_\chi^{\text{an}}]) \in H^1(Y, \mathbb{Q}(\chi))$, where $\mathbb{Q}(\chi)$ is the finite extension of $\mathbb{Q}$ obtained by adjoining the values of χ.

Let $L \supset \mathbb{Q}(\chi)$ be any finite extension of $\mathbb{Q}_p$. Then similar considerations show that $G_{\mathbb{Q}}$ acts on the corresponding class $c_{\chi,\text{ét}} = \iota_{\text{ét}}(c_\chi) \in H^1_{\text{ét}}(\overline{Y}, L)$ via $\bar{\chi}\epsilon^{-1}$, that is, $Lc_{\chi,\text{ét}} \simeq L(\bar{\chi}\epsilon^{-1})$ as $L[G_{\mathbb{Q}}]$-modules. We then obtain an extension $\mathcal{E}^{\text{mod}}_{\chi,\text{ét}}$ as a subquotient of the relative cohomology group $H^1_{\text{ét}}(\overline{Y}, \overline{W}, L)$ analogously to $\mathcal{E}_{\chi,\text{ét}}$, where now $W = C_\infty$. Let $\lambda^{\text{mod}}_{\chi,\text{ét}} : H^0_{\text{ét}}(\overline{W}, L) \to L$ be the $L[G_{\mathbb{Q}}]$-homomorphism such that

$$\lambda^{\text{mod}}_{\chi,\text{ét}}((c_P)_{P \in C_\infty}) = c_{\left[\begin{smallmatrix}1\\N\end{smallmatrix}\right]} - c_{\left[\begin{smallmatrix}a\\N\end{smallmatrix}\right]}$$

for some fixed a with $(a, N) = 1$, $a \not\equiv \pm 1 \bmod N$. Note that $\lambda^{\text{mod}}_{\chi,\text{ét}}$ is trivial on the image of $H^0_{\text{ét}}(\overline{Y}, L)$. The extension $\mathcal{E}^{\text{mod}}_{\chi,\text{ét}}$ is the pullback/pushout

$$\begin{tikzcd}
\frac{H^0_{\text{ét}}(\overline{W},L)}{\operatorname{im} H^0_{\text{ét}}(\overline{Y},L)} \arrow[r, hook] \arrow[d, two heads, "\lambda^{\text{mod}}_{\chi,\text{ét}}"] & H^1_{\text{ét}}(\overline{Y}, \overline{W}, L) \arrow[r, two heads] \arrow[d, dashed] & H^1_{\text{ét}}(\overline{Y}, L) \\
L \arrow[r, hook] & \mathcal{E}^{\text{mod}}_{\chi,\text{ét}} \arrow[r, two heads] & Lc_{\chi,\text{ét}}, \arrow[u, hook]
\end{tikzcd}$$

with the dashed arrow denoting a subquotient.

We analyze the extension

$$\mathcal{E}^{\text{mod}}_{\chi,\text{ét}} \in \operatorname{Ext}^1_{L[G_{\mathbb{Q}}]}(L, Lc_{\chi,\text{ét}}) = \operatorname{Ext}^1_{L[G_{\mathbb{Q}}]}(L, L(\bar{\chi}\epsilon^{-1}))$$

just as we did $\mathcal{E}_{\chi,\text{ét}}$ in Section 5. Suppose — again for simplicity — that χ is valued in $\mathbb{Q}_p$ (so we may take $L = \mathbb{Q}_p$) and $p \nmid N$. Then $D_{\text{crys}}(\mathbb{Q}_p c_{\chi,\text{ét}}) = D_{\text{dR}}(\mathbb{Q}_p c_{\chi,\text{ét}}) = \mathbb{Q}_p(c_{\chi,\text{ét}} \otimes \tau(\bar{\chi})\underline{t})$, and it is easy to see — by comparing residues at cusps — that $\iota_{\text{dR},p}([\omega^{\text{alg}}_\chi]) = c_{\chi,\text{ét}} \otimes \frac{1}{\tau(\chi)}\underline{t}$, where $\omega^{\text{alg}}_\chi = 2\pi i \omega_\chi \in H^0(\Omega^1_{X/\mathbb{Q}_p}(\log C_0))$. Specifically, ϕ_p acts on $[\omega^{\text{alg}}_\chi]$ as multiplication by $\chi(p)p$ and we seek to understand whether $(1 - \bar{\chi}(p)p^{-1}\phi_p)[\omega^{\text{alg}}_\chi]_W \in H^1_{\text{dR}}((Y, W)/\mathbb{Q}_p)$ is the image of something nontrivial in $H^0_{\text{dR}}(W/\mathbb{Q}_p)$ that is nonzero under $\lambda^{\text{mod}}_{\chi,\text{ét}}$. We now replace the passage to Monsky–Washnitzer cohomology with restriction to the rigid cohomology of the ordinary locus of X (the rigid analytic subspace of points corresponding to elliptic curves with ordinary reduction at p) and also with partial compact support in W. This moves the calculation into the realm of overconvergent p-adic modular forms, just as passage to MW cohomology moved the calculation to the realm of overconvergent functions on the affinoid Y_{an} in Sections 5.5 and 5.6. The action of ϕ_p on a p-adic modular form $f(q) \in \mathbb{Q}_p[\![q]\!]$ of weight 2 is just $f(q) \mapsto pf(q^p)$, and the differential on p-adic modular functions is just the p-adic Maass–Shimura operator $\theta = q\frac{d}{dq}$. In particular, we want to find an overconvergent p-adic modular function $\eta(q)$ (a form of weight 0) such that $\theta\eta = G_\chi(q) - \bar{\chi}(p)G_\chi(q^p)$. Then $(1 - \bar{\chi}(p)p^{-1}\phi_p)[\omega^{\text{alg}}_\chi]_W$ is the image of $\eta|_W \in H^0_{\text{dR}}(W/\mathbb{Q}_p)$, and so we want to know whether $\lambda^{\text{mod}}_{\chi,\text{ét}}(\eta|_W) \neq 0$. It is easy to identify η from the q-expansion of $G_\chi(q) - \bar{\chi}(p)G_\chi(q^p)$: $\eta = E^{\text{ord}}_{\bar{\chi},0}$, the p-ordinary weight-0 Eisenstein series with

q-expansion

$$E^{\mathrm{ord}}_{\bar{\chi},0}(q) = \tfrac{1}{2}L_p(1,\bar{\chi}) + \sum_{n=1}^{\infty}\biggl(\sum_{\substack{d \bmod n \\ p\nmid d}} \bar{\chi}(d)d^{-1}\biggr)q^n.$$

The existence of such an Eisenstein series is an easy consequence of Katz's Eisenstein measure (which also provides a proof of the existence of the p-adic L-function $L_p(s,\bar{\chi})$ as a p-adic measure — see [Serre 1973] and [Katz 1975]). It follows that

$$\lambda^{\mathrm{mod}}_{\chi,\text{ét}}(\eta|W) = c_{\left[\frac{1}{N}\right]}(E^{\mathrm{ord}}_{\chi,0}) - c_{\left[\frac{a}{N}\right]}(E^{\mathrm{ord}}_{\chi,0}) = \tfrac{1}{2}(1-\chi(a))L_p(1,\bar{\chi}).$$

For a satisfying $\chi(a) \neq 1$, this shows that $\mathcal{E}^{\mathrm{mod}}_{\chi,\text{ét}}|_{G_{\mathbb{Q}_p}}$ is nonsplit if $L_p(1,\bar{\chi}) \neq 0$. One can associate with $\mathcal{E}^{\mathrm{mod}}_{\chi,\text{ét}}$ a cohomology class $z^{\mathrm{mod}}_{\chi} \in H^1_f(\mathbb{Q},\mathbb{Q}_p(\chi\epsilon))$ by tensoring the extension over $\mathbb{Q}_p$ with $\mathbb{Q}_p(\chi\epsilon)$, just as we associated z_χ with $\mathcal{E}_{\chi,\text{ét}}$. Then unwinding the preceding analysis as in Section 5.7 yields

$$\widetilde{\log}_{\mathrm{BK}}(\mathrm{loc}_p(z^{\mathrm{mod}}_{\chi})) = \tfrac{1}{2}(1-\chi(a))L_p(1,\bar{\chi})\cdot(1\otimes\underline{\zeta}\otimes\underline{t}^{-1}) \in D_{\mathrm{crys}}(\mathbb{Q}_p(\chi\epsilon)).$$

6.1.1. Remark. We conclude with a few remarks:

(1) Unlike for z_χ, which was constructed from the cohomology of $\mathbb{P}^1\setminus\{\infty,1\}$, this computation of the Bloch–Kato logarithm of z^{mod}_{χ} does not rely on a formula for the special value $L_p(1,\bar{\chi})$ in terms of p-adic logs of cyclotomic units, but instead comes naturally via the value of a constant term of a p-adic Eisenstein series, and it is via the latter that Serre and Katz (*re-*)*constructed* the p-adic L-function [Serre 1973; Katz 1975]. The construction of z^{mod}_{χ} (via $\mathcal{E}^{\mathrm{mod}}_{\chi,\text{ét}}$) can be viewed as a cohomological expression of the Serre–Katz construction. Our next construction of cohomology classes — for Hecke characters of imaginary quadratic fields — lends itself to a similar interpretation.

(2) The class z^{mod}_{χ} can be extended to an Euler system

$$\bigl\{z^{\mathrm{mod}}_{\chi,M} \in H^1(\mathbb{Q}[\mu_M],\mathbb{Z}_p(\chi\epsilon)) : (M,N)=1\bigr\}.$$

The classes $z^{\mathrm{mod}}_{\chi,M}$ are just the cohomology classes associated with extensions constructed via pullback/pushforward from simple, natural variations on the Eisenstein classes $\omega_\chi = G_\chi(\tau)d\tau$. However, unlike for z_χ (and $z_{\chi,M}$), the construction described above does not immediately imply that the class z^{mod}_{χ} (or $z^{\mathrm{mod}}_{\chi,M}$) belongs to $H^1(\mathbb{Q},\mathbb{Z}_p(\chi\epsilon))$. This can be shown, though, via a more careful use of the comparison isomorphisms of p-adic Hodge theory: Assuming $p\nmid N$, we can work with smooth integral models of X, Y, $Z = X\setminus Y$, and $W = C_\infty$ over $\mathbb{Z}_p$. Then $\omega_\chi \in H^0(\Omega^1_{X/\mathbb{Z}_p}(\log Z))$ and $\iota_{\mathrm{dR},p} : H^1_{\mathrm{dR}}(Y/\mathbb{Z}_p) \xrightarrow{\sim} (H^1_{\text{ét}}(\overline{Y},\mathbb{Z}_p)\otimes_{\mathbb{Z}_p} A_{\mathrm{crys}})^{G_{\mathbb{Q}_p}}$, where $A_{\mathrm{crys}} \subset B_{\mathrm{crys}}$ is the usual integral crystalline ring. As $\underline{t}$ is not divisible by a nonunit of $\mathbb{Z}_p^{\mathrm{ur}}$ in A_{crys}, the relation $\iota_{\mathrm{dR}}([\omega^{\mathrm{alg}}_{\chi}]) = c_{\chi,\text{ét}}\otimes\frac{1}{\tau(\chi)}\underline{t}$ then implies that

$c_{\chi,\text{ét}} \in H^1_{\text{ét}}(\overline{Y}, \mathbb{Z}_p)$. A variation of this argument, similar to [Faltings 2005, §10], can be used to handle the case when $p \mid N$ (and also $z^{\text{mod}}_{\chi,M}$ when $p \mid M$).

(3) As mention in Section 3, a very similar construction of extensions can be found in Harder's unpublished work [2023]. Essentially the same construction can also be found in unpublished work of Romyar Sharifi and Preston Wake. However, neither detect nonsplitting without reference to a comparison with the extension classes defined by modular units.

6.2. *Hecke characters.* Let ℓ be a prime, and let $X = X_0(\ell)$ and $Y_0(\ell)$ be the usual modular curves for the congruence subgroup $\Gamma_0(\ell)$, which we view as smooth curves over $\mathbb{Q}$ via the usual canonical models. The cusps $C = X \setminus Y$ consists of two points, usually denoted ∞ and 0 and both defined over $\mathbb{Q}$. The unique holomorphic Eisenstein series E of weight 2, level ℓ, and trivial character, which has q-expansion

$$E(\tau) = \frac{(1-\ell)\zeta(-1)}{2} + \sum_{n=1}^{\infty}\Bigg(\sum_{\substack{d\,|\,n\\ \ell\nmid d}} d\Bigg)q^n, \quad q = e^{2\pi i\tau},$$

defines a class $\omega_E = E(\tau)d\tau \in H^0(\Omega^1_{X/\mathbb{C}}(\log C))$ and $c_E = \iota_{\text{dR}}[\omega_E] \in H^1(Y, \mathbb{C})$ actually belongs to $H^1(Y, \mathbb{Q})$. So $c_{E,\text{ét}} = \iota_{\text{ét}}(c_E) \in H^1_{\text{ét}}(\overline{Y}, \mathbb{Q}_p)$. The action of $G_{\mathbb{Q}}$ on $c_{E,\text{ét}}$ is via ϵ^{-1}. That is, $\mathbb{Q}_p c_{E,\text{ét}} \simeq \mathbb{Q}_p(-1)$ as a $\mathbb{Q}_p[G_{\mathbb{Q}}]$-module.

Let K be an imaginary quadratic field with ring of integers $\mathcal{O}$. Let

$$R = \left\{\begin{pmatrix} a & b \\ c & d\end{pmatrix} \in M_2(\mathbb{Z}) : \ell \mid c\right\}$$

be the usual Eichler order of level ℓ (so $(R \otimes \widehat{\mathbb{Z}}) \cap \mathrm{GL}_2(\mathbb{Q})^+ = \Gamma_0(\ell)$). Fix an embedding $K \hookrightarrow M_2(\mathbb{Q})$ such that $R \cap K = \mathcal{O}$. Let $\tau_0 \in \mathbb{H}$ be such that its stabilizer in $\mathrm{GL}_2(\mathbb{Q})^+$ is $K^\times$. Then

$$W = \big\{[\tau_0, x] \in Y(\mathbb{C}) = \mathrm{GL}_2(\mathbb{Q})^+ \setminus [\mathbb{H} \times \mathrm{GL}_2(\mathbb{A}_f)/(R \otimes \widehat{\mathbb{Z}})^\times] : x \in (K \otimes \mathbb{A}_f)^\times\big\}$$

is a collection of CM points on Y. It is in bijection with the class group of K. The set W is defined over K and each point in W is defined over the Hilbert class field H of K. The action of G_K on W is described via CM theory: Let $\mathrm{Art}_K : K^\times \setminus (K \otimes \mathbb{A}_f)^\times \twoheadrightarrow G_K^{\text{ab}}$ be Artin map of class field theory, with geometric normalizations. If $\sigma \in G_K$ is such that the image of $\sigma \in G_K^{\text{ab}}$ is $\mathrm{Art}_K(z)$ then $\sigma \cdot [\tau_0, x] = [\tau, zx]$.

We view W as a K-subvariety of Y. Let $\psi : K^\times \setminus (K \otimes \mathbb{A}_f)^\times/(\mathcal{O} \otimes \widehat{\mathbb{Z}})^\times \to \mathbb{C}^\times$ be a character of the class group of K. We also view this as a character of G_K via the projection to $\mathrm{Gal}(H/K)$ and the Artin map. Suppose — for simplicity — that ψ takes values in $\mathbb{Q}_p$. Then $\lambda^K_{\psi,\text{ét}} : H^0_{\text{ét}}(\overline{W}, \mathbb{Q}_p) \twoheadrightarrow \mathbb{Q}_p(\psi)$, $\lambda^K_{\psi,\text{ét}}((c_w)_{w\in W}) = \sum_{w=[\tau_0,x]\in W} \psi(x)c_w$, is a G_K-equivariant map. Here $\mathbb{Q}_p(\psi)$ is just $\mathbb{Q}_p$ with

$\sigma \in G_K$ acting via multiplication by $\psi(\sigma)$. The usual pull-back/push-forward construction then yields an extension $\mathcal{E}^K_{\psi,\text{ét}}$:

$$\begin{tikzcd}
\frac{H^0_{\text{ét}}(\overline{W},\mathbb{Q}_p)}{\mathrm{im} H^0_{\text{ét}}(\overline{Y},\mathbb{Q}_p)} \arrow[r, hook] \arrow[d, two heads, "\lambda^K_{\psi,\text{ét}}"] & H^1_{\text{ét}}(\overline{Y}, \overline{W}, \mathbb{Q}_p) \arrow[r, two heads] \arrow[d, dashed] & H^1_{\text{ét}}(\overline{Y}, \mathbb{Q}_p) \\
\mathbb{Q}_p(\psi) \arrow[r, hook] & \mathcal{E}^K_{\psi,\text{ét}} \arrow[r, two heads] & \mathbb{Q}_p c_{E,\text{ét}}, \arrow[u, hook]
\end{tikzcd}$$

which defines a class in $\mathrm{Ext}^1_{\mathbb{Q}_p[G_K]}(\mathbb{Q}_p(\psi), \mathbb{Q}_p c_{E,\text{ét}}) = \mathrm{Ext}^1_{\mathbb{Q}_p[G_K]}(\mathbb{Q}_p(\psi), \mathbb{Q}_p(-1))$. And associated with this is a class $z^K_\psi \in H^1(K, \mathbb{Q}_p(\psi\epsilon))$.

Suppose p splits in K: $p = v\bar{v}$. The Bloch–Kato logarithm of $\mathrm{loc}_v(z_\psi) \in H^1_f(K_v, \mathbb{Q}_p(\psi\epsilon))$ can be computed following the same method employed for $\mathrm{loc}_p(z^{\mathrm{mod}}_\chi)$. The upshot is that $\widetilde{\log}_{\mathrm{BK}}(\mathrm{loc}_v(z^K_\psi))$ is a multiple of a natural basis of $D_{\mathrm{crys}}(\mathbb{Q}_p(\psi\epsilon))$, with that multiple being expressed as

$$\sum_{w=[\tau_0,x]\in W} \psi(x) E^{\mathrm{ord}}_0(w), \tag{6.2.a}$$

where E^{ord}_0 is the p-ordinary weight-0 Eisenstein series with q-expansion

$$E^{\mathrm{ord}}_0(q) = (1-\ell^{-1})\zeta_p(1) + \sum_{n=1}^{\infty}\Bigg(\sum_{\substack{d\,|\,n \\ p \nmid d \\ \ell \nmid n/d}} d^{-1}\Bigg) q^n.$$

Note that $\theta E^{\mathrm{ord}}_0 = E(q) - E(q^p)$ which is identified with $(1 - p^{-1}\phi_p)[\omega_E]$ in the rigid cohomology of the ordinary locus of Y, so the expression (6.2.a) is just $\lambda_{\psi,\mathrm{dR}}(E^{\mathrm{ord}}_0|W)$. Via Katz's construction of the p-adic L-function of $\bar{\psi}$ relative to the choice of v [1975], the expression (6.2.a) can be seen to be a simple multiple of the value at $s = 1$ of the p-adic L-function. That is, the Bloch–Kato logarithm of $\mathrm{loc}_v(z^K_\psi)$ is naturally expressed as a value of a p-adic L-function for $\bar{\psi}$.

6.2.1. Remark. Just as for z^{mod}_χ, the class z^K_ψ can be extended to an Euler system for $\mathbb{Z}_p(\psi\epsilon)$ over K in the sense of Rubin [2000]. This involves varying W over CM points defined over ring (and even ray) class extensions as well as varying the Eisenstein class. In this way, one can recover/reconstruct the Euler system for ψ over K previously defined by Rubin [1991] using elliptic units along with its connection with Katz's two-variable p-adic L-function.

Acknowledgements

This work was supported by the Simons Investigator Grant #376203 from the Simons Foundation and the National Science Foundation Grant DMS-1901985. It is a pleasure to thank the referees for comments on an earlier draft and suggestions for making it more reader-friendly. Particular thanks go to Fernando Trejos Suárez for his careful reading of the earlier version and for his detailed comments. It is also a pleasure to recognize the interest, encouragement, and helpful remarks from Adebisi Agboola, Henri Darmon, and Preston Wake.

References

[Beilinson 1984] A. A. Beilinson, "Higher regulators and values of L-functions", *Itogi Nauki i Tekhniki. Ser. Sovrem. Probl. Mat. Nov. Dostizh.* **24** (1984), 181–238. In Russian; translated in *J. Math. Sci. New York* **30**:2 (1985), 2036–2070. MR

[Berger 2004] L. Berger, "An introduction to the theory of p-adic representations", pp. 255–292 in *Geometric aspects of Dwork theory*, vol. 1, edited by A. Adolphson et al., de Gruyter, Berlin, 2004. MR Zbl

[Berger 2013] L. Berger, "On p-adic Galois representations", pp. 3–19 in *Elliptic curves*, *Hilbert modular forms and Galois deformations*, edited by H. Darmon et al., Springer, 2013. MR Zbl

[Besser 2012] A. Besser, "Heidelberg lectures on Coleman integration", pp. 3–52 in *The arithmetic of fundamental groups—PIA* 2010, edited by J. Stix, Contrib. Math. Comput. Sci. **2**, Springer, 2012. MR Zbl

[Bloch and Kato 1990] S. Bloch and K. Kato, "L-functions and Tamagawa numbers of motives", pp. 333–400 in *The Grothendieck Festschrift*, *I*, edited by P. Cartier et al., Progr. Math. **86**, Birkhäuser, Boston, 1990. MR Zbl

[Brown 2014] F. Brown, "Motivic periods and $\mathbb{P}^1 \setminus \{0, 1, \infty\}$", pp. 295–318 in *Proceedings of the International Congress of Mathematicians—Seoul* 2014, *II*, edited by S. Y. Jang et al., Kyung Moon Sa, Seoul, 2014. MR Zbl

[Carlson 1980] J. A. Carlson, "Extensions of mixed Hodge structures", pp. 107–127 in *Journées de Géometrie Algébrique* (Angers, France, 1979), edited by A. Beauville, Sijthoff & Noordhoff, Alphen aan den Rijn, 1980. MR Zbl

[Carlson and Hain 1989] J. A. Carlson and R. M. Hain, "Extensions of variations of mixed Hodge structure", pp. 39–65 in *Théorie de Hodge* (Luminy, France, 1987), Astérisque **179-180**, Société Mathématique de France, Paris, 1989. MR Zbl

[Conrad and Brinon 2009] B. Conrad and O. Brinon, "CMI summer school notes on p-adic Hodge theory (preliminary version)", lecture notes, 2009, available at https://virtualmath1.stanford.edu/~conrad/.

[Deligne 1971a] P. Deligne, "Théorie de Hodge. I", pp. 425–430 in *Actes du Congrès International des Mathématiciens*, *Tome* 1 (Nice, France, 1970), Gauthier-Villars, Paris, 1971. MR Zbl

[Deligne 1971b] P. Deligne, "Théorie de Hodge, II", *Inst. Hautes Études Sci. Publ. Math.* **40** (1971), 5–57. MR Zbl

[Deligne 1974] P. Deligne, "Théorie de Hodge, III", *Inst. Hautes Études Sci. Publ. Math.* **44** (1974), 5–77. MR Zbl

[Deligne 1989] P. Deligne, "Le groupe fondamental de la droite projective moins trois points", pp. 79–297 in *Galois groups over* $\mathbb{Q}$ (Berkeley, CA, 1987), edited by Y. Ihara et al., Math. Sci. Res. Inst. Publ. **16**, Springer, 1989. MR Zbl

[Deligne and Goncharov 2005] P. Deligne and A. B. Goncharov, "Groupes fondamentaux motiviques de Tate mixte", *Ann. Sci. École Norm. Sup.* (4) **38**:1 (2005), 1–56. MR Zbl

[Esnault and Viehweg 1992] H. Esnault and E. Viehweg, *Lectures on vanishing theorems*, DMV Seminar **20**, Birkhäuser, Basel, 1992. MR Zbl

[Faltings 2005] G. Faltings, "Arithmetic Eisenstein classes on the Siegel space: some computations", pp. 133–166 in *Number fields and function fields—two parallel worlds*, edited by G. van der Geer et al., Progr. Math. **239**, Birkhäuser, Boston, 2005. MR Zbl

[Fontaine and Perrin-Riou 1994] J.-M. Fontaine and B. Perrin-Riou, "Autour des conjectures de Bloch et Kato: cohomologie galoisienne et valeurs de fonctions L", pp. 599–706 in *Motives* (Seattle, WA, 1991), edited by U. Jannsen et al., Proc. Sympos. Pure Math. **55, Part 1**, Amer. Math. Soc., Providence, RI, 1994. MR Zbl

[Harder 2023] G. Harder, "Eisenstein cohomology and the construction of mixed motives", unpublished notes, 2023, available at https://www.math.uni-bonn.de/people/harder/.

[Huber 2020] A. Huber, "Galois theory of periods", *Münster J. Math.* **13**:2 (2020), 573–596. MR Zbl

[Huber and Müller-Stach 2017] A. Huber and S. Müller-Stach, *Periods and Nori motives*, Ergebnisse der Math. (3) **65**, Springer, 2017. MR Zbl

[Katz 1975] N. M. Katz, "p-adic L-functions via moduli of elliptic curves", pp. 479–506 in *Algebraic geometry* (Arcata, CA, 1974), edited by R. Hartshorne, Proc. Sympos. Pure Math. **29**, Amer. Math. Soc., Providence, RI, 1975. MR Zbl

[Kedlaya 2008] K. S. Kedlaya, "p-adic cohomology: from theory to practice", pp. 175–203 in *p-adic geometry*, edited by D. Savitt and D. S. Thakur, Univ. Lecture Ser. **45**, Amer. Math. Soc., Providence, RI, 2008. MR Zbl

[Nekovář 1994] J. Nekovář, "Beilinson's conjectures", pp. 537–570 in *Motives* (Seattle, WA, 1991), edited by U. Jannsen et al., Proc. Sympos. Pure Math. **55, Part 1**, Amer. Math. Soc., Providence, RI, 1994. MR Zbl

[Peters and Steenbrink 2008] C. A. M. Peters and J. H. M. Steenbrink, *Mixed Hodge structures*, Ergebnisse der Math. (3) **52**, Springer, 2008. MR Zbl

[van der Put 1986] M. van der Put, "The cohomology of Monsky and Washnitzer", pp. 33–59 in *Introductions aux cohomologies p-adiques* (Luminy, France, 1984), edited by D. Barsky and P. Robba, Mém. Soc. Math. France **23**, Soc. Math. France, Paris, 1986. MR Zbl

[Rubin 1991] K. Rubin, "The 'main conjectures' of Iwasawa theory for imaginary quadratic fields", *Invent. Math.* **103**:1 (1991), 25–68. MR Zbl

[Rubin 2000] K. Rubin, *Euler systems*, Annals of Mathematics Studies **147**, Princeton University Press, 2000. MR Zbl

[Sangiovanni-Vincentelli and Skinner $\geq$ 2024a] M. A. Sangiovanni-Vincentelli and C. Skinner, "An Euler system for the symmetric square of a modular form", In preparation.

[Sangiovanni-Vincentelli and Skinner $\geq$ 2024b] M. A. Sangiovanni-Vincentelli and C. Skinner, "Euler systems for unitary groups", In preparation.

[Serre 1973] J.-P. Serre, "Formes modulaires et fonctions zêta p-adiques", pp. 191–268 in *Modular functions of one variable, III* (Antwerp, 1972), edited by W. Kuyk and J.-P. Serre, Lecture Notes in Math. **350**, Springer, 1973. MR Zbl

[Shang et al. $\geq$ 2024] Z. Shang, C. Skinner, and F. Trejos Suárez, "Eisenstein classes on modular curves", In preparation.

[Washington 1997] L. C. Washington, *Introduction to cyclotomic fields*, 2nd ed., Graduate Texts in Mathematics **83**, Springer, 1997. MR Zbl

Received 12 Jan 2023. Revised 27 Jun 2024.

CHRISTOPHER SKINNER:

cmcls@princeton.edu
Department of Mathematics, Princeton University, Princeton, NJ, United States

msp

Essential Number Theory
Vol. 3, No. 1, 2024
https://doi.org/10.2140/ent.2024.3.101

Subconvexity implies effective quantum unique ergodicity for Hecke–Maaß cusp forms on $\mathrm{SL}_2(\mathbb{Z}) \backslash \mathrm{SL}_2(\mathbb{R})$

Ankit Bisain, Peter Humphries,
Andrei Mandelshtam, Noah Walsh and Xun Wang

It is a folklore result in arithmetic quantum chaos that quantum unique ergodicity on the modular surface with an effective rate of convergence follows from subconvex bounds for certain triple product L-functions. The physical space manifestation of this result, namely the equidistribution of mass of Hecke–Maaß cusp forms, was proven to follow from subconvexity by Watson, whereas the phase space manifestation of quantum unique ergodicity has only previously appeared in the literature for Eisenstein series via work of Jakobson. We detail the analogous phase space result for Hecke–Maaß cusp forms. The proof relies on the Watson–Ichino triple product formula together with a careful analysis of certain archimedean integrals of Whittaker functions.

1. Introduction

Quantum ergodicity, in its most general sense, originates from the study of quantum chaos. Loosely speaking, quantum ergodicity for a Riemannian manifold is the notion that almost all eigenfunctions of the Laplace–Beltrami operator equidistribute in the large eigenvalue limit. The foundational *quantum ergodicity theorem* due to Shnirelman [1974] proves quantum ergodicity for a compact Riemannian manifold with ergodic geodesic flow. In the language of quantum chaos, this can be seen as going from chaotic classical mechanics of a system to equidistribution of energy eigenstates of the system.

We begin with a brief introduction to the general case of quantum ergodicity. We then introduce *arithmetic quantum chaos*, which will be the focus for the remainder of this paper. In the setting of arithmetic quantum chaos, notions such as quantum ergodicity are studied on manifolds with arithmetic structure, giving the eigenfunctions additional structure that is not present in the generic case. For surveys of the generic case of quantum ergodicity, see [Anantharaman 2010; De Bièvre 2001; Dyatlov 2022; Hassell 2011; Nonnenmacher 2013; Zelditch 2006; 2010; 2019], while for surveys on arithmetic quantum chaos, see [Marklof 2006; Sarnak 2011].

MSC2020: primary 11F12; secondary 11F66, 11F67, 58J51, 81Q50.
Keywords: quantum chaos, quantum unique ergodicity, subconvexity.

1A. ***Quantum ergodicity.***

1A1. *Classical dynamics.* Let (M, g) be a smooth compact oriented n-dimensional Riemannian manifold. The cotangent bundle T^*M of the manifold M consists of points (x, ξ) with $x \in M$ and $\xi \in T_x^*M$, the space of tangent covectors at x. Associated to each point $x \in M$ and tangent vector $v \in T_xM$ is a unique geodesic $\gamma : \mathbb{R} \to M$ for which $\gamma(0) = x$ and $\gamma'(0) = v$, which gives rise to the geodesic flow $G^t(x, v) := (\gamma(t), \gamma'(t))$ on the tangent bundle TM and the corresponding geodesic flow $G_t(x, \xi)$ on the cotangent bundle T^*M by identifying tangent covectors with the corresponding tangent vectors. Equivalently, the geodesic flow $G_t(x, \xi) = (x(t), \xi(t))$ is the Hamiltonian flow, obtained as solutions to the Hamiltonian equations

$$\frac{d}{dt} x_j(t) = \frac{\partial}{\partial \xi_j} H(x(t), \xi(t)), \qquad \frac{d}{dt} \xi_j(t) = -\frac{\partial}{\partial \xi_j} H(x(t), \xi(t)), \qquad j \in \{1, \dots, n\},$$

of the Hamiltonian

$$H(x, \xi) := \sum_{j,k=1}^{n} g^{jk}(x) \xi_j \xi_k$$

on T^*M, where $g_{jk}(x) = g\big(\frac{\partial}{\partial x_j}, \frac{\partial}{\partial x_k}\big)$ denotes the Riemannian metric g, while $g^{jk}(x)$ denotes the entries of the inverse matrix. This geodesic flow preserves the cosphere bundle S^*M, which consists of points $(x, \xi) \in T^*M$ with ξ of unit length.

One can think of the manifold M as being *physical space* that encodes the position of a point particle on M, while the cosphere bundle S^*M is *phase space* and encodes both the position and momentum of a point particle. The geodesic flow on S^*M then encodes the position and momentum of a point particle over time and describes the *classical dynamics* on M.

The metric g induces probability measures μ and ω on M and S^*M respectively. The latter is called the *Liouville measure* and is invariant under the geodesic flow. The geodesic flow on S^*M is said to be *ergodic* if for almost every $(x, \xi) \in S^*M$, the geodesic flow $G_t(x, \xi)$ equidistributes on S^*M with respect to the Liouville measure. Ergodic geodesic flow demonstrates that the classical dynamics on M are *chaotic*.

1A2. *Quantum dynamics.* These notions for classical dynamics on M have quantum dynamical counterparts. Point particles are replaced by quantum particles with wave functions $\psi : M \times \mathbb{R} \to \mathbb{C}$. In place of S^*M, the space of states is instead wave functions ψ that are square-integrable with respect to the volume measure μ on M; with ψ L^2-normalized, the probability density $|\psi(x, t)|^2 \, d\mu(x)$ then encodes the probability that a quantum particle is located in a region at a given time t.

The classical dynamics G_t are replaced by the quantum dynamics given by the evolution of a wave function over time as governed by Schrödinger's equation

$$i\frac{\partial\psi}{\partial t} = -\Delta\psi.$$

Here

$$\Delta := -\frac{1}{\sqrt{|\det g|}}\sum_{j,k=1}^{n}\frac{\partial}{\partial x_j}g^{jk}(x)\sqrt{|\det g|}\frac{\partial}{\partial x_k}$$

denotes the Laplace–Beltrami operator on M, which we refer to as the Laplacian for the sake of brevity. This is a second order scalar linear partial differential operator, defined from the Riemannian metric g, that commutes with isometries of M; moreover, it is, up to scalar multiplication, the unique nontrivial such scalar linear partial differential operator of minimal order. The quantization of the Hamiltonian H is the Laplacian Δ, while the quantization of the geodesic flow G_t acting on S^*M is the quantum flow acting on $L^2(M)$ given by the Schrödinger propagator $U^t := e^{-it\Delta}$.

Stationary states are L^2-normalized wave functions $\psi(x,t)$ that are separable, so that $\psi(x,t) = \varphi(x)\nu(t)$ for some functions $\varphi : M \to \mathbb{C}$ and $\nu : \mathbb{R} \to \mathbb{C}$, and are such that the probability densities $|\psi(x,t)|^2\,\mathrm{d}\mu(x)$ are independent of time, so that $\varphi \in L^2(M)$ and $\nu(t) = e^{-i\lambda t}$ for some $\lambda \in \mathbb{R}$. The functions φ correspond to solutions of the eigenvalue problem

$$\Delta\varphi = \lambda\varphi;$$

that is, they are Laplacian eigenfunctions. For each eigenvalue, the corresponding eigenspace of Laplacian eigenfunctions is finite-dimensional, so that one can choose an orthonormal basis of each eigenspace. The union of these orthonormal bases then forms an orthonormal basis of the whole space $L^2(M)$, and the corresponding collection of eigenvalues forms a countable discrete set of nonnegative real numbers. We denote the set of Laplacian eigenfunctions by $(\varphi_j)_{j\geq 1}$ and the corresponding Laplacian eigenvalues by $(\lambda_j)_{j\geq 1}$.

Associated to each Laplacian eigenfunction φ_j is its *microlocal lift* ω_j, alternatively known as a *Wigner distribution*. The microlocal lift is a distribution on S^*M of the measure corresponding to φ_j on the cosphere bundle S^*M, as defined in [Dyatlov 2022, (2)]; it should be thought of as measuring the average value in phase space S^*M of an observable $a \in C^\infty(S^*M)$ for a quantum particle with wave function φ_j. When acting on observables $a \in C^\infty(S^*M)$ that descend to functions on M, so that $a(x,\xi)$ is constant in ξ the microlocal lift is simply the distribution $a(x,\xi)|\varphi_j(x)|^2\,\mathrm{d}\mu(x)$ on M.

1A3. *Quantum ergodicity.* In [Lazutkin 1993, Addendum], Shnirelman proved that if the geodesic flow on S^*M is ergodic with respect to the Liouville measure ω, there exists a subsequence $(\varphi_{j_k})_{k\geq 1}$ of the sequence of Laplacian eigenfunctions $(\varphi_j)_{j\geq 1}$ of density 1 (in the sense that $\#\{\lambda_{j_k} \leq \lambda\}/\#\{\lambda_k \leq \lambda\} \to 1$ as $\lambda \to \infty$) such that for all smooth functions a on M,

$$\lim_{k\to\infty} \int_M a(x)|\varphi_{j_k}(x)|^2 \, \mathrm{d}\mu(x) = \int_M a(x) \, \mathrm{d}\mu(x).$$

That is, a density 1 subsequence of the eigenfunctions *equidistributes in physical space*. Shnirelman in fact proved the stronger statement that a density 1 subsequence $(\omega_{j_k})_{k\geq 1}$ of the sequence of microlocal lifts *equidistributes in phase space* in the sense that it approaches the Liouville measure on S^*M. That is, for any smooth function a on S^*M,

$$\lim_{k\to\infty} \int_{S^*M} a(x,\xi) \, \mathrm{d}\omega_{j_k}(x,\xi) = \int_{S^*M} a(x,\xi) \, \mathrm{d}\omega(x,\xi).$$

This property is known as *quantum ergodicity*. An outline of a proof of the quantum ergodicity theorem similar to Shnirelman's original proof can be found in [Dyatlov 2022, Section 2]. Shnirelman's proof was first announced in [Shnirelman 1974] and independent proofs were obtained by Zelditch [1987] and Colin de Verdière [1985]. Quantum ergodicity demonstrates that the quantum dynamics on M are *chaotic*. The fact that quantum ergodicity follows from the assumption that the geodesic flow on S^*M is ergodic can be viewed as showing that chaotic classical dynamics imply chaotic quantum dynamics.

Quantum *unique* ergodicity (QUE) in physical space is the property that $(\varphi_j)_{j\geq 1}$ satisfies

$$\lim_{j\to\infty} \int_M a(x)|\varphi_j(x)|^2 \, \mathrm{d}\mu(x) = \int_M a(x) \, \mathrm{d}\mu(x)$$

for all smooth functions a on M. Equivalently, QUE in physical space is the property that the whole sequence of eigenfunctions equidistributes in physical space M. The notion of QUE has a natural generalization to phase space S^*M: quantum unique ergodicity in phase space refers to the property of $(\varphi_j)_{j\geq 1}$ satisfying

$$\lim_{j\to\infty} \int_{S^*M} a(x,\xi) \, \mathrm{d}\omega_j(x,\xi) = \int_{S^*M} a(x,\xi) \, \mathrm{d}\omega(x,\xi) \tag{1}$$

for all smooth functions a on S^*M. Henceforth, QUE will refer to quantum unique ergodicity on phase space unless otherwise noted.

It was established by Hassell [2010, Theorem 1] that there exist compact Riemannian manifolds for which the geodesic flow is ergodic and yet not all eigenfunctions equidistribute. Namely, Hassell showed that QUE does not hold for a large family of stadium domains.[1] However, in many cases, it is still believed that QUE should

[1] For manifolds with boundary, the geodesic flow is replaced by the *billiard flow*, where trajectories bounce off of the boundary.

hold. In particular, it was conjectured by Rudnick and Sarnak [1994, Conjecture] that QUE holds when (M, g) is a compact hyperbolic surface or more generally a negatively curved compact manifold.

1B. *Quantum unique ergodicity for arithmetic surfaces.* For most hyperbolic surfaces, QUE is far from proven. However, this conjecture is better understood in the case where (M, g) is an arithmetic hyperbolic surface.

Let $\mathbb{H} := \{z = x + iy \in \mathbb{C} : y \in \mathbb{R}_+\}$ denote the upper half-plane with area measure $\mathrm{d}\mu(z) := \frac{\mathrm{d}x\,\mathrm{d}y}{y^2}$ and Laplacian $\Delta := -y^2\big(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\big)$ coming from the standard hyperbolic metric $\mathrm{d}s^2 := \frac{\mathrm{d}x^2+\mathrm{d}y^2}{y^2}$. Recall that $\mathrm{SL}_2(\mathbb{R})$ acts transitively on $\mathbb{H}$ via Möbius transformations, namely

$$g \cdot z := \frac{az+b}{cz+d} \quad \text{for } g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{R}) \text{ and } z \in \mathbb{H}.$$

If $\Gamma \subset \mathrm{SL}_2(\mathbb{R})$ is an arithmetic subgroup (in the sense of [Katok 1992, Chapter 5]), such as a congruence subgroup of $\mathrm{SL}_2(\mathbb{Z})$, the quotient space $\Gamma\backslash\mathbb{H}$ is an *arithmetic hyperbolic surface*. These surfaces are not necessarily compact, but have finite area, allowing the necessary notions to be defined. In particular, $\Gamma\backslash\mathbb{H}$ has finite area (with respect to $\mathrm{d}\mu$) given by $\frac{\pi}{3}[\mathrm{SL}_2(\mathbb{Z}) : \Gamma]$ when Γ is a finite-index subgroup of the modular group $\mathrm{SL}_2(\mathbb{Z})$. Note that in general, the area measure $d\mu(z)$ is *not* a probability measure on $\Gamma\backslash\mathbb{H}$, which differs from the normalization of the probability measure μ on M in Section 1A.

The study of QUE on arithmetic surfaces is aided via the presence of *Hecke operators* (see (11) below). The Hecke operators on a given arithmetic hyperbolic surface are a sequence $T_1, T_2, \ldots$ of self-adjoint operators on the space of square-integrable functions on the surface; they are a nonarchimedean analogue of the Laplace–Beltrami operator. It is known that Hecke operators commute with each other and with the hyperbolic Laplacian Δ. We may therefore simultaneously diagonalize the space of Maaß cusp forms (nonconstant Laplacian eigenfunctions occurring in the discrete spectrum of the Laplacian) with respect to the Hecke operators, obtaining a basis of *Hecke–Maaß cusp forms*, which are simultaneous eigenfunctions of both the Laplacian and of all the Hecke operators. Due to the additional structure given from the Hecke operators, stronger results regarding QUE are known for such Hecke eigenbases.

Henceforth, we focus on the case where $\Gamma = \mathrm{SL}_2(\mathbb{Z})$ and $M = \mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}$ is the modular surface. This surface is not compact, as it has a cusp at $i\infty$. Its cosphere bundle S^*M may be identified with the quotient space $\mathrm{SL}_2(\mathbb{Z})\backslash \mathrm{SL}_2(\mathbb{R})$, while the microlocal lift ω_j of a Laplacian eigenfunction can be explicitly expressed in terms of linear combinations of raised and lowered Laplacian eigenfunctions, as we explicate further in Section 3B. Its Laplacian eigenfunctions can be split into two classes. There is a discrete spectrum, which, besides constant functions, arises from nonconstant Laplacian eigenfunctions φ_j called *Maaß cusp forms* corresponding to a nondecreasing sequence of positive eigenvalues λ_j; the *cuspidality condition* is precisely the condition that

$$\int_0^1 \varphi_j(x+iy)\,\mathrm{d}x = 0$$

for all $y \in \mathbb{R}_+$. Because M is noncompact, there is also a continuous spectrum, with eigenfunctions coming from real-analytic Eisenstein series $E\bigl(z, \frac{1}{2}+it\bigr)$ with eigenvalues $\frac{1}{4}+t^2$. We discuss Hecke–Maaß cusp forms and Eisenstein series in further detail in Section 3B; see [Duke et al. 2002, Section 4]).

It is a seminal result of Lindenstrauss [2006, Theorem 1.4] that on a (possibly noncompact) arithmetic hyperbolic surface, for a Hecke eigenbasis, any limit (in the weak-* topology) of a subsequence of the measures ω_j is a nonnegative multiple of the Liouville measure ω. When the surface is compact, this limit must be the Liouville measure itself, proving QUE for compact arithmetic hyperbolic surfaces; see [Sarnak 2011, Section 3] for more discussion of the relevant work and progress in the arithmetic case.

On the (noncompact) modular surface, equidistribution for the continuous spectrum was established in physical space by Luo and Sarnak [1995, Theorem 1.1], and later in phase space by Jakobson [1994, Theorem 1]. Since the modular surface is noncompact, the work of Lindenstrauss does not establish QUE for this surface, as there is possibility of mass escaping to the cusp. This possibility was eliminated by Soundararajan [2010], establishing QUE for Hecke–Maaß cusp forms on the modular surface. However, this resolution of QUE for Hecke–Maaß cusp forms leaves unresolved the problem of determining the *rate* of equidistribution.

Jakobson [1997, Theorem 2] proves that the measures ω_j converge to ω in an averaged sense with an effective rate of averaged equidistribution. Precisely, Jakobson proves that if a is an element of the space $C_{c,K}^{\infty}(S^*M)$ consisting of finite linear combinations of smooth compactly supported functions of even weight (as described in (9) below), then

$$\sum_{\lambda_j \le \lambda} \left| \int_{S^*M} a(z,\theta)\,\mathrm{d}\omega_j(z,\theta) - \int_{S^*M} a(z,\theta)\,\mathrm{d}\omega(z,\theta) \right|^2 \ll_{a,\varepsilon} \lambda^{1/2+\varepsilon}. \tag{2}$$

As Weyl's law implies that the number of eigenvalues below λ is asymptotic to $\frac{\lambda}{12}$ [Risager 2004, Theorem 2], this gives an averaged bound of $\lambda^{-1/2+\varepsilon}$ on each summand. This bound generalized an earlier result of Luo and Sarnak [1995, Theorem 1.2], which essentially gave the analogous average bound in physical space. Luo and Sarnak also remark that the best possible individual bound for each summand in (2) is of size $\lambda_j^{-1/2}$. To see why this is true, we recall that it was established by Sarnak and Zhao [2019, Theorem 1.1] that

$$\sum_{\lambda_j \le \lambda} \left| \int_{S^*M} a(z,\theta)\, \mathrm{d}\omega_j(z,\theta) - \int_{S^*M} a(z,\theta)\, \mathrm{d}\omega(z,\theta) \right|^2 \sim Q(a,a)\lambda^{1/2},$$

where $Q(a,b)$ is a fixed sesquilinear form on

$$C_{c,K}^{\infty}(\mathrm{SL}_2(\mathbb{Z})\backslash \mathrm{SL}_2(\mathbb{R})) \times C_{c,K}^{\infty}(\mathrm{SL}_2(\mathbb{Z})\backslash \mathrm{SL}_2(\mathbb{R})).$$

It follows that if

$$\max_{\lambda_j \le \lambda} \left| \int_{S^*M} a(z,\theta)\, \mathrm{d}\omega_j(z,\theta) - \int_{S^*M} a(z,\theta)\, \mathrm{d}\omega(z,\theta) \right| \le C$$

for some nonnegative constant C, then

$$\sum_{\lambda_j \le \lambda} \left| \int_{S^*M} a(z,\theta)\, \mathrm{d}\omega_j(z,\theta) - \int_{S^*M} a(z,\theta)\, \mathrm{d}\omega(z,\theta) \right|^2 \le C^2\left(\frac{\lambda}{12} + o(\lambda)\right),$$

which are contradictory statements unless $C \gg \lambda^{-1/4}$.

1C. *Results.* Our goal is to prove bounds for the *individual* terms

$$\int_{S^*M} a(z,\theta)\, \mathrm{d}\omega_j(z,\theta) - \int_{S^*M} a(z,\theta)\, \mathrm{d}\omega(z,\theta).$$

These bounds are contingent on bounds for certain L-functions. Watson [2002, Theorem 3] establishes the following formula for integrals of products of Hecke–Maaß cusp forms φ_j, whose precise definitions we give in Section 3B: there exists a nonnegative absolute constant C such that

$$\left| \int_M \varphi_{j_1}(z)\varphi_{j_2}(z)\varphi_{j_3}(z)\, \mathrm{d}\mu(z) \right|^2 = C \frac{\Lambda\left(\frac{1}{2}, \varphi_{j_1} \otimes \varphi_{j_2} \otimes \varphi_{j_3}\right)}{\Lambda(1, \mathrm{ad}\,\varphi_{j_1})\Lambda(1, \mathrm{ad}\,\varphi_{j_2})\Lambda(1, \mathrm{ad}\,\varphi_{j_3})}.$$

Here the terms on the right-hand side are completed L-functions whose definitions are given in Section 4B. The Lindelöf hypothesis for such L-functions (itself a consequence of the generalized Riemann hypothesis) would then imply sufficiently strong upper bounds in order to prove the uniform version of Luo and Sarnak's physical space result [1995, Theorem 1.2]. In particular, for any $a \in C_c^{\infty}(M)$, we would have that

$$\int_M a(z)|\varphi_j(z)|^2\, \mathrm{d}\mu(z) - \int_M a(z)\, \mathrm{d}\mu(z) \ll_{a,\varepsilon} \lambda_j^{-1/4+\varepsilon}$$

under the assumption of the conjectural bound $L\left(\frac{1}{2}, \varphi_{j_1} \otimes \varphi_{j_1} \otimes \varphi_{j_3}\right) \ll_{\varphi_{j_3},\varepsilon} \lambda_{j_1}^{\varepsilon}$; see [Watson 2002, Corollary 1] and [Young 2016, Proposition 1.5]. More generally, any effective subconvex bound of the form $L\left(\frac{1}{2}, \varphi_{j_1} \otimes \varphi_{j_1} \otimes \varphi_{j_3}\right) \ll_{\varphi_{j_3}} \lambda_{j_1}^{1/2-2\delta}$ would provide the above statement with weaker error term of the form $O_a(\lambda_j^{-\delta} \log \lambda_j)$.

In this paper, we prove the strengthening of this physical space statement to phase space.

Theorem 1.1. *Suppose that there exist constants $\delta > 0$ and $A > 0$ such that for any Hecke–Maaß cusp forms φ_1, φ_2 with Laplacian eigenvalues λ_1, λ_2, any $t \in \mathbb{R}$, and any holomorphic Hecke cusp form F, we have the subconvex bounds*

$$L\big(\tfrac{1}{2}, \operatorname{ad}\varphi_1 \otimes \varphi_2\big) \ll \lambda_1^{1/2-2\delta} \lambda_2^{A}, \tag{3}$$

$$L\big(\tfrac{1}{2} + it, \operatorname{ad}\varphi_1\big) \ll \lambda_1^{1/4-\delta} (1+|t|)^{A}, \tag{4}$$

$$L\big(\tfrac{1}{2}, \operatorname{ad}\varphi_1 \otimes F\big) \ll_F \lambda_1^{1/2-2\delta}. \tag{5}$$

Then for any $a \in C_{c,K}^{\infty}(S^*M)$, we have that

$$\int_{S^*M} a(z,\theta)\,\mathrm{d}\omega_j(z,\theta) - \int_{S^*M} a(z,\theta)\,\mathrm{d}\omega(z,\theta) \ll_a \lambda_j^{-\delta} \log \lambda_j. \tag{6}$$

In particular, assuming the generalized Lindelöf hypothesis, we have that

$$\int_{S^*M} a(z,\theta)\,\mathrm{d}\omega_j(z,\theta) - \int_{S^*M} a(z,\theta)\,\mathrm{d}\omega(z,\theta) \ll_{a,\varepsilon} \lambda_j^{-1/4+\varepsilon}.$$

Remark 1.2. The method of proof yields explicit dependence on a in these error terms in terms of Sobolev norms of a; see (64).

The subconvex bounds (3), (4), and (5) are *hypotheses* in Theorem 1.1: it is not yet known that these subconvex bounds hold. While subconvex bounds are known for certain other families of L-functions (see, for example, [Michel and Venkatesh 2010; Nelson 2023]), unconditional proofs of the desired subconvex bounds (3), (4), and (5) currently remain elusive. A chief obstacle towards proving these bounds is the so-called *conductor dropping phenomenon*, as discussed in [Khan and Young 2023].

Theorem 1.1 is folklore (see, for example, [Sarnak and Zhao 2019, page 1156]), though no detailed proof exists in the literature. The method of proof is known to experts; the analogue of QUE for Bianchi manifolds (i.e., arithmetic quotients of $\mathbb{H}^3 = \mathrm{SL}_2(\mathbb{C})/\mathrm{SU}(2)$), for example, has been shown by Marshall to follow from subconvexity for triple product L-functions [Marshall 2014, Theorem 3], and the proof that we give for the modular surface is by the same general strategy. To explicate all the details, one needs the full strength of the Watson–Ichino triple product formula as in [Watson 2002, Theorem 3] and [Ichino 2008, Theorem 1.1]. Coupling this with a lemma of Michel and Venkatesh [Michel and Venkatesh 2010, Lemma 3.4.2] (compare [Sarnak and Zhao 2019, Lemma 5]), we show that certain triple products of automorphic forms on $\mathrm{SL}_2(\mathbb{Z})\backslash \mathrm{SL}_2(\mathbb{R})$ can be expressed in terms of a product of central values of L-functions and certain archimedean integrals of Whittaker functions; the latter can in turn be related to gamma functions and hypergeometric functions.

Finally, we take this opportunity to observe that Jakobson's treatment of QUE for Eisenstein series in [Jakobson 1994] is incomplete; in particular, the case where the test function is a shifted holomorphic or antiholomorphic Hecke cusp form is missing. We supply the omitted computations in Section 5.

1D. ***Friedrichs symmetrization.*** We end the discussion of our results by explaining how our results are valid not only for the Wigner distribution ω_j, which need not be a positive distribution, but also for the *Friedrichs symmetrization* ω_j^F defined in (7) below, which *is* a positive distribution. The microlocal lifts ω_j of Hecke–Maaß cusp forms on the modular surface that we work with in this paper are the Wigner distributions given by

$$\mathrm{d}\omega_j(z,\theta) := \varphi_j(z)\overline{u_j(z,\theta)}\,\mathrm{d}\omega(z,\theta), \qquad u_j(z,\theta) := \frac{3}{\pi}\sum_{k=-\infty}^{\infty} \varphi_{j,k}(z)e^{2ki\theta},$$

as defined in [Zelditch 1991, (1.18)]. Here the convergence is in distribution and $\mathrm{d}\omega$ is the (unnormalized) Liouville measure, given by $\frac{\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}\theta}{2\pi y^2}$ on $\mathrm{SL}_2(\mathbb{Z})\backslash \mathrm{SL}_2(\mathbb{R}) = S^*M$, where we identify $g \in \mathrm{SL}_2(\mathbb{R})$ with $(x, y, \theta) \in \mathbb{R} \times \mathbb{R}_+ \times [0, 2\pi)$ via the Iwasawa decomposition (see (8) below). The functions $\varphi_{j,k}$ are the L^2-normalized shifted Hecke–Maaß forms of weight $2k$ obtained from φ_j by raising or lowering operators, as defined in Section 3B; for their Fourier expansions, see Section 4A.

We recall that a positive distribution T on a normed space V over $\mathbb{C}$ is a bounded linear functional $T : V \to \mathbb{C}$ such that $T(v) \geq 0$ for all $v \in V$. In general, the Wigner distribution $\mathrm{d}\omega_j$ need not be a positive distribution on $C_c^\infty(\mathrm{SL}_2(\mathbb{Z})\backslash \mathrm{SL}_2(\mathbb{R}))$. To convert $\mathrm{d}\omega_j$ into a positive distribution, we define for $a \in C_c^\infty(\mathrm{SL}_2(\mathbb{Z})\backslash \mathrm{SL}_2(\mathbb{R}))$ the pairing

$$\begin{aligned}\langle a, \mathrm{d}\omega_j\rangle &= \int_{S^*M} a(z,\theta)\,\mathrm{d}\omega_j(z,\theta)\\ &:= \lim_{K\to\infty}\int_{\mathrm{SL}_2(\mathbb{Z})\backslash \mathrm{SL}_2(\mathbb{R})} a(z,\theta)\varphi_j(z)\sum_{k=-K}^{K}\overline{\varphi_{j,k}(z)e^{2ki\theta}}\,\mathrm{d}\omega(z,\theta).\end{aligned}$$

We now define a new distribution $\mathrm{d}\omega_j^F$, the *Friedrichs symmetrization* of $\mathrm{d}\omega_j$, via

$$\langle a, \mathrm{d}\omega_j^F\rangle := \langle a^F, \mathrm{d}\omega_j\rangle, \tag{7}$$

where the function $a^F \in C_c^\infty(\mathrm{SL}_2(\mathbb{Z})\backslash \mathrm{SL}_2(\mathbb{R}))$ is the Friedrichs symmetrization of a; for its explicit construction, see [Zelditch 1987, Proposition 2.3]. In particular, it was established in [loc. cit., Proposition 2.3] that $\mathrm{d}\omega_j^F$ is a positive distribution,[2] while it was established in [Zelditch 1991, Proposition 3.8] that

$$\langle a, \mathrm{d}\omega_j^F\rangle - \langle a, \mathrm{d}\omega_j\rangle \ll_{a,\varepsilon} \lambda_j^{-1/2+\varepsilon}.$$

Combined with Theorem 1.1, we see that in specific scenarios where one needs to deal with positive distributions, it suffices to work with the Wigner distribution $\mathrm{d}\omega_j$.

2. Proof outline

On a broad scale, our proof strategy follows the proof of equidistribution of Eisenstein series in phase space from [Jakobson 1994], which we now outline. We will also make reference to a few objects that we have not yet defined; namely, we use (x, y, θ) coordinates on S^*M given by (8), L-functions that we explain in Section 4B, and various types of functions on M all defined in Section 3B.

In our paper, we extend the probability measure $|\varphi_j|^2\,\mathrm{d}\mu$ to its microlocal lift $\mathrm{d}\omega_j$ on S^*M for a Hecke–Maaß cusp form φ_j with Laplacian eigenvalue λ_j. The work of Jakobson [1994] solves a similar problem: Jakobson proves the analogous result for the extension of the Radon measure $\left|E\left(\,\cdot\,, \tfrac{1}{2}+it\right)\right|^2 \mathrm{d}\mu$ to its microlocal lift $\mathrm{d}\mu_t$. Jakobson's method for bounding integrals of the form $\int a\,\mathrm{d}\mu_t$ is to consider only functions a appearing in an orthonormal basis of $L^2(S^*M)$. Namely, Jakobson computes the integral for constant functions, shifted Hecke–Maaß cusp forms, shifted holomorphic or antiholomorphic Hecke cusp forms,[3] and weighted Eisenstein series. He then bounds $\int a\,\mathrm{d}\mu_t$ for general smooth, compactly supported a on S^*M by approximating them using this basis.

To bound $\int a\,\mathrm{d}\mu_t$, Jakobson uses the coordinates (x, y, θ) on S^*M and proceeds to integrate over θ, which reduces the problem to computing integrals over M. These integrals can readily be evaluated using the key fact that they involve Eisenstein series. An Eisenstein series can be written by a sum over $\Gamma_\infty\backslash \mathrm{SL}_2(\mathbb{Z})$, where $\Gamma_\infty \subset \mathrm{SL}_2(\mathbb{Z})$ is the stabilizer of the cusp $i\infty$, in such a way that the integral can be unfolded to one over the fundamental domain $\{x + iy \in \mathbb{H} : x \in [0, 1]\}$ for $\Gamma_\infty\backslash\mathbb{H}$. Jakobson then inserts the Fourier–Whittaker expansion of each function in the integrand and subsequently directly evaluates the integral over $x \in [0, 1]$. One is left with an expression involving central values of L-functions related to the

[2]Lindenstrauss [Lindenstrauss 2001, Corollary 3.2] constructs an alternate positive distribution that has a similar effect: for each $N \in \mathbb{N}$, Lindenstrauss defines the positive distribution $\mathrm{d}\omega_j^N(z,\theta) := \frac{3}{\pi}\frac{1}{2N+1}\left|\sum_{k=-N}^{N} \varphi_{j,k}(z)e^{2ki\theta}\right|^2 \mathrm{d}\omega(z,\theta)$. For $N \sim \lambda_j^{1/4}$, this satisfies $\langle a, \mathrm{d}\omega_j^N\rangle - \langle a, \mathrm{d}\omega_j\rangle \ll_{a,\varepsilon} \lambda_j^{-1/4+\varepsilon}$.

[3]As mentioned previously, Jakobson only treats unshifted holomorphic Hecke cusp forms and neglects to deal with the more general case of shifted holomorphic or antiholomorphic Hecke cusp forms. We complete Jakobson's proof by dealing with this untreated general case in Section 5.

test functions and an integral over $y \in \mathbb{R}_+$ of Whittaker functions. This remaining integral can be expressed in terms of hypergeometric functions and subsequently bounded using Stirling's formula.

Our paper follows a similar reduction of integrals, using the same orthonormal basis. In particular, we must show that the constant term contributes the main term in Theorem 1.1, while the contribution from integrating against shifted Hecke–Maaß cusp forms, shifted holomorphic or antiholomorphic Hecke cusp forms, and shifted Eisenstein series are $O(\lambda_j^{-\delta} \log \lambda_j)$ as $j \to \infty$. We now outline how we evaluate each type of integral:

- The constant case is trivial, and contributes to the main term in Theorem 1.1.
- The weighted Eisenstein series case can be computed with an unfolding technique analogous to the previously discussed computations in [Jakobson 1994]. Computing this integral gives a product of a central value of an L-function and an expression involving gamma functions and hypergeometric functions.
- For the remaining two cases, namely shifted Hecke–Maaß cusp forms and shifted holomorphic or antiholomorphic Hecke cusp forms, the unfolding trick does not apply to the integrals of interest since they do not involve an Eisenstein series. Instead, we use the *Watson–Ichino triple product formula* [Ichino 2008; Watson 2002]. This formula allows us to write the square of the absolute value of the integral as a product of a central value of an L-function and the square of the absolute value of an integral of Whittaker functions. The latter integral can again be explicitly computed to obtain an expression in terms of hypergeometric functions.

We then bound all hypergeometric functions using Stirling's formula, while we invoke our assumption of subconvexity to bound central values of L-functions, which yields Theorem 1.1.

3. Preliminaries

3A. ***Differential operators on*** $\mathbf{SL_2(\mathbb{Z})\backslash SL_2(\mathbb{R})}$ ***and*** $\mathbf{SL_2(\mathbb{Z})\backslash \mathbb{H}}$. We begin by describing differential operators acting on $\mathrm{SL}_2(\mathbb{Z})\backslash \mathrm{SL}_2(\mathbb{R})$ and $\mathrm{SL}_2(\mathbb{Z})\backslash \mathrm{SL}_2(\mathbb{R})$. Useful references for these include [Bump 1997, Chapter 2; Iwaniec 2002, Chapter 1; Lang 1985; Roelcke 1966].

In coordinates $z = x + iy \in \mathbb{H}$, the Laplacian on $\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}$ is given by $\Delta := -y^2\big(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\big)$, and the area measure is given by $\mathrm{d}\mu(z) := \frac{\mathrm{d}x\,\mathrm{d}y}{y^2}$, giving this space volume $\frac{\pi}{3}$. The unnormalized Liouville measure on the unit cotangent bundle $S^*M = \mathrm{SL}_2(\mathbb{Z})\backslash \mathrm{SL}_2(\mathbb{R})$ is given by $\mathrm{d}\omega(z,\theta) := \frac{\mathrm{d}\mu(z)\,\mathrm{d}\theta}{2\pi}$, which also gives this space volume $\frac{\pi}{3}$. Here we identify points on S^*M with points on $\mathrm{SL}_2(\mathbb{Z})\backslash \mathrm{SL}_2(\mathbb{R})$

using the Iwasawa decomposition

$$g = \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} \begin{pmatrix} y^{1/2} & 0 \\ 0 & y^{-1/2} \end{pmatrix} \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \tag{8}$$

for elements $g \in \mathrm{SL}_2(\mathbb{R})$, where $x \in \mathbb{R}$, $y \in \mathbb{R}_+$, and $\theta \in [0, 2\pi)$.

A function $\Phi : \mathrm{SL}_2(\mathbb{Z}) \backslash \mathrm{SL}_2(\mathbb{R}) \to \mathbb{C}$ is of weight $2k$ for some $k \in \mathbb{Z}$ if it satisfies

$$\Phi(z, \theta + \theta') = e^{2ki\theta'} \Phi(z, \theta) \tag{9}$$

for all $z \in \mathbb{H}$ and $\theta, \theta' \in \mathbb{R}$. We have an inner product on $\mathrm{SL}_2(\mathbb{Z}) \backslash \mathrm{SL}_2(\mathbb{R})$ defined by

$$\langle \Phi_1, \Phi_2 \rangle := \int_{\mathrm{SL}_2(\mathbb{Z}) \backslash \mathrm{SL}_2(\mathbb{R})} \Phi_1(z, \theta) \overline{\Phi_2(z, \theta)} \, \mathrm{d}\omega(z, \theta)$$

for $\Phi_1, \Phi_2 \in L^2(\mathrm{SL}_2(\mathbb{Z}) \backslash \mathrm{SL}_2(\mathbb{R}))$.

Similarly, a function $f : \mathbb{H} \to \mathbb{C}$ satisfying the automorphy condition

$$f\left(\frac{az + b}{cz + d}\right) = \left(\frac{cz + d}{|cz + d|}\right)^{2k} f(z) \tag{10}$$

for all $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$ is said to be of weight $2k$. A weight $2k$ function $f : \mathbb{H} \to \mathbb{C}$ lifts to a weight $2k$ function $\Phi : \mathrm{SL}_2(\mathbb{Z}) \backslash \mathrm{SL}_2(\mathbb{R}) \to \mathbb{C}$ via the map $f \mapsto \Phi$ given by $\Phi(z, \theta) := f(z) e^{2ki\theta}$. Equivalently, we define for $g = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{R})$ and $z \in \mathbb{H}$ the j-factor

$$j_g(z) := \frac{cz + d}{|cz + d|},$$

so that for $g \in \mathrm{SL}_2(\mathbb{R})$ given by (8), we have that $\Phi(z, \theta) := j_g(i)^{-k} f(g \cdot i)$. We have an inner product on weight $2k$ functions $f_1, f_2 : \mathbb{H} \to \mathbb{C}$ defined by

$$\langle f_1, f_2 \rangle := \int_{\mathrm{SL}_2(\mathbb{Z}) \backslash \mathbb{H}} f_1(z) \overline{f_2(z)} \, \mathrm{d}\mu(z).$$

The $\mathrm{SL}_2(\mathbb{R})$-invariant extension of Δ from functions on $\mathbb{H}$ to functions on $\mathrm{SL}_2(\mathbb{R})$ is given by the Casimir operator

$$\Omega := -y^2 \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right) + y \frac{\partial^2}{\partial x \partial \theta}.$$

We also have raising and lowering operators

$$R := e^{2i\theta} iy \left(\frac{\partial}{\partial x} - i \frac{\partial}{\partial y}\right) - e^{2i\theta} \frac{i}{2} \frac{\partial}{\partial \theta}, \quad L := -e^{-2i\theta} iy \left(\frac{\partial}{\partial x} + i \frac{\partial}{\partial y}\right) + e^{-2i\theta} \frac{i}{2} \frac{\partial}{\partial \theta}.$$

The operators Ω, R, L are initially defined on the space $C^\infty(\mathrm{SL}_2(\mathbb{Z}) \backslash \mathrm{SL}_2(\mathbb{R}))$ of smooth functions. The raising and lowering operators R and L map weight $2k$ eigenfunctions of Ω to weight $2k + 2$ and $2k - 2$ eigenfunctions of Ω respectively.

On the space $C^\infty(\mathrm{SL}_2(\mathbb{Z})\backslash \mathrm{SL}_2(\mathbb{R})) \cap L^2(\mathrm{SL}_2(\mathbb{Z})\backslash \mathrm{SL}_2(\mathbb{R}))$, these operators are such that $-L$ is adjoint to R, while Ω is self-adjoint, so that for all smooth square-integrable functions $\Phi_1, \Phi_2 : \mathrm{SL}_2(\mathbb{Z})\backslash \mathrm{SL}_2(\mathbb{R}) \to \mathbb{C}$,

$$\langle R\Phi_1, \Phi_2\rangle = -\langle \Phi_1, L\Phi_2\rangle, \quad \langle \Omega\Phi_1, \Phi_2\rangle = \langle \Phi_1, \Omega\Phi_2\rangle.$$

The Casimir operator Ω admits a canonical self-adjoint extension, the *Friedrichs extension*, to $L^2(\mathrm{SL}_2(\mathbb{Z})\backslash \mathrm{SL}_2(\mathbb{R}))$; see [Iwaniec 2002, Theorem A.3].

These operators descend to $\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}$. Considering the action of Ω on weight $2k$ functions on $\mathbb{H}$, we have the corresponding weight $2k$ Laplacian on $\mathbb{H}$ given by

$$\Delta_{2k} := \Delta + 2kiy\frac{\partial}{\partial x},$$

which preserves the weight of a weight $2k$ function on $\mathbb{H}$. Similarly, R and L become the raising and lowering operators

$$R_{2k} := iy\left(\frac{\partial}{\partial x} - i\frac{\partial}{\partial y}\right) + k, \quad L_{2k} := -iy\left(\frac{\partial}{\partial x} + i\frac{\partial}{\partial y}\right) - k.$$

The raising operator R_{2k} maps weight $2k$ functions to weight $2k+2$ functions, whereas the lowering operator L_{2k} maps weight $2k$ functions to weight $2k-2$ functions. In particular, the raising and lowering operators map eigenfunctions of Δ_{2k} to eigenfunctions of Δ_{2k+2} and Δ_{2k-2} respectively. The inner product on weight $2k$ functions on $\mathbb{H}$ is such that $-L_{2k+2}$ is adjoint to R_{2k}, so that

$$\langle R_{2k} f_1, f_2\rangle = -\langle f_1, L_{2k+2} f_2\rangle$$

for all weight $2k$ square-integrable functions $f_1 : \mathbb{H} \to \mathbb{C}$ and weight $2k+2$ square-integrable functions $f_2 : \mathbb{H} \to \mathbb{C}$.

3B. ***Eigenfunctions of the Laplacian.*** Next, we describe the eigenfunctions of Δ_{2k} on $\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}$. Useful references for these include [Bump 1997, Chapter 2; Duke et al. 2002, Section 4; Roelcke 1966].

For any $k \in \mathbb{Z}$, there are up to four classes of eigenfunctions of Δ_{2k} of weight $2k$. Each of these is an eigenfunction of the n-th Hecke operator T_n for each $n \in \mathbb{N}$, where T_n acts on functions $f : \mathbb{H} \to \mathbb{C}$ via

$$(T_n f)(z) := \frac{1}{\sqrt{n}} \sum_{ad=n} \sum_{b=1}^{d} f\left(\frac{az+b}{d}\right). \tag{11}$$

Each of these eigenfunctions of Δ_{2k} also lifts to a function on $\mathrm{SL}_2(\mathbb{Z})\backslash \mathrm{SL}_2(\mathbb{R})$ that is an eigenfunction of Ω:

• When $k = 0$, we have constant functions.

• When $k \geq 0$, we have shifted Maaß cusp forms of weight $2k$ given by $R_{2k-2} \cdots R_0\varphi_j$, where φ_j is a Hecke–Maaß cusp form of weight 0 with j-th Laplacian eigenvalue λ_j (ordered by size). Similarly, when $k \leq 0$ we have forms of weight $2k$ given by $L_{2k+2}L_{2k+4} \cdots L_0\varphi_j$. Any weight 0 form φ_j can be written as a sum of an even part and an odd part with the same Laplacian and Hecke eigenvalues, so we may additionally assume that φ_j is either even, so that $\varphi_j(-\bar{z}) = \varphi_j(z)$, or odd, so that $\varphi_j(-\bar{z}) = -\varphi_j(z)$. We let $\kappa_j \in \{0, 1\}$ be such that κ_j is 0 if φ_j is even and κ_j is 1 if φ_j is odd; the *parity* of φ_j is then defined to be $\epsilon_j = (-1)^{\kappa_j}$, so that

$$\varphi_j(-\bar{z}) = (-1)^{\kappa_j}\varphi_j(z) = \epsilon_j\varphi_j(z). \tag{12}$$

The *spectral parameter* $r_j \in [0, \infty) \cup i(0, \frac{1}{2})$ satisfies $\lambda_j = \frac{1}{4} + r_j^2$; since the Selberg eigenvalue conjecture is known for $\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}$, r_j must be real and positive (with the smallest spectral parameter being $r_1 \approx 9.534$). Once L^2-normalized with respect to the measure $d\mu$ on $\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}$, the eigenfunctions φ_j yield probability measures $d\mu_j = |\varphi_j|^2\, d\mu$ on $\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}$. The corresponding L^2-normalized shifted Hecke–Maaß cusp forms of weight $2k$ are given by

$$\varphi_{j,k} := \begin{cases} \dfrac{\Gamma\left(\frac{1}{2}+ir_j\right)}{\Gamma\left(\frac{1}{2}+k+ir_j\right)} R_{2k-2} \cdots R_2 R_0 \varphi_j & \text{for } k \geq 0, \\ \dfrac{\Gamma\left(\frac{1}{2}+ir_j\right)}{\Gamma\left(\frac{1}{2}-k+ir_j\right)} L_{2k+2} \cdots L_{-2} L_0 \varphi_j & \text{for } k \leq 0, \end{cases}$$

where the L^2-normalization of each $\varphi_{j,k}$ follows from the fact that $-L_{2k+2}$ is adjoint to R_{2k}; see [Duke et al. 2002, Corollary 4.4]. The associated lift to $\mathrm{SL}_2(\mathbb{Z})\backslash \mathrm{SL}_2(\mathbb{R})$ is the function $\Phi_{j,k}(z, \theta) := \varphi_{j,k}(z)e^{2ki\theta}$, which is an eigenfunction of the Casimir operator Ω with eigenvalue λ_j.

• When $\ell \geq 1$, let F be a holomorphic Hecke cusp form of weight 2ℓ; there are finitely many such cusp forms, and we denote the set of such holomorphic Hecke cusp forms by $\mathcal{H}_\ell$. We define a corresponding weight 2ℓ function $f(z) = y^\ell F(z)$, which is automorphic of weight 2ℓ, so that it satisfies the automorphy condition (10) with $k = \ell$. When $k \geq \ell$, we have shifted holomorphic Hecke cusp forms of weight $2k$ given by $R_{2k-2}R_{2k-4} \cdots R_{2\ell} f$. Similarly, when $k \leq -\ell$ we have the shifted antiholomorphic Hecke cusp form of weight $2k$ given by $L_{2k+2}L_{2k+4} \cdots L_{-2\ell}\bar{f}$. Note that $L_{2\ell} f = R_{-2\ell}\bar{f} = 0$, so that there are no nonzero shifted cusp forms of weight $2k$ with $-\ell < k < \ell$. If f is L^2-normalized with respect to the measure $d\mu$ on $\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}$, then the corresponding L^2-normalized shifted holomorphic or antiholomorphic Hecke cusp forms of weight $2k$ are given by

$$f_k := \begin{cases} \sqrt{\dfrac{\Gamma(2\ell)}{\Gamma(k+\ell)\Gamma(k-\ell+1)}} R_{2k-2} \cdots R_{2\ell} f & \text{for } k \geq \ell, \\ \sqrt{\dfrac{\Gamma(2\ell)}{\Gamma(-k+\ell)\Gamma(-k-\ell+1)}} L_{2k+2} \cdots L_{-2\ell} f & \text{for } k \leq -\ell. \end{cases}$$

Once more, the L^2-normalization of each f_k follows from the fact that $-L_{2k+2}$ is adjoint to R_{2k}; see [Duke et al. 2002, Corollary 4.4]. The associated lift to $\mathrm{SL}_2(\mathbb{Z})\backslash \mathrm{SL}_2(\mathbb{R})$ is the function $\Psi_{F,k}(z,\theta) := f_k(z)e^{2ki\theta}$, which is an eigenfunction of the Casimir operator Ω with eigenvalue $\ell(1-\ell)$.

- We have the Eisenstein series of weight $2k$, which is defined by

$$E_{2k}(z,s) := \sum_{\gamma\in\Gamma_\infty\backslash \mathrm{SL}_2(\mathbb{Z})} j_\gamma(z)^{-2k}\Im(\gamma z)^s, \tag{13}$$

where

$$j_\gamma(z) := \frac{cz+d}{|cz+d|} \text{ for } \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

This series converges absolutely when $\Re(s) > 1$, and can be holomorphically extended to the line $\Re(s) = \frac{1}{2}$. For $s = \frac{1}{2}+it$, $E_{2k}\bigl(z, \frac{1}{2}+it\bigr)$ is an eigenfunction of Δ_{2k} with eigenvalue $\frac{1}{4}+t^2$. Letting $E(z,s) = E_0(z,s)$, we note that

$$E_{2k}\bigl(z, \tfrac{1}{2}+it\bigr) = \begin{cases} \dfrac{\Gamma\bigl(\frac{1}{2}+it\bigr)}{\Gamma\bigl(\frac{1}{2}+k+it\bigr)} R_{2k-2}\cdots R_2R_0E\bigl(z, \frac{1}{2}+it\bigr) & \text{for } k \geq 0, \\ \dfrac{\Gamma\bigl(\frac{1}{2}+it\bigr)}{\Gamma\bigl(\frac{1}{2}-k+it\bigr)} L_{2k+2}\cdots L_{-2}L_0E\bigl(z, \frac{1}{2}+it\bigr) & \text{for } k \leq 0. \end{cases}$$

The associated lift to $\mathrm{SL}_2(\mathbb{Z})\backslash \mathrm{SL}_2(\mathbb{R})$ is

$$\widetilde{E}_{2k}\bigl(z,\theta, \tfrac{1}{2}+it\bigr) := E_{2k}\bigl(z, \tfrac{1}{2}+it\bigr)e^{2ki\theta},$$

which is an eigenfunction of the Casimir operator Ω with eigenvalue $\frac{1}{4}+t^2$.

These Laplacian eigenfunctions satisfy orthonormality relations: we have that

$$\begin{gathered} \langle \Phi_{j,k}, 1\rangle = \langle \Psi_{F,k}, 1\rangle = 0, \\ \bigl\langle \Phi_{j,k_1}, \widetilde{E}_{2k_2}\bigl(\cdot,\cdot, \tfrac{1}{2}+it\bigr)\bigr\rangle = \bigl\langle \Psi_{F,k_1}, \widetilde{E}_{2k_2}\bigl(\cdot,\cdot, \tfrac{1}{2}+it\bigr)\bigr\rangle = 0, \\ \langle \Phi_{j,k_1}, \Psi_{F,k_2}\rangle = 0, \\ \langle \Phi_{j_1,k_1}, \Phi_{j_2,k_2}\rangle = \begin{cases} 1 & \text{if } j_1 = j_2 \text{ and } k_1 = k_2, \\ 0 & \text{otherwise,} \end{cases} \\ \langle \Psi_{F_1,k_1}, \Psi_{F_2,k_2}\rangle = \begin{cases} 1 & \text{if } F_1 = F_2 \text{ and } k_1 = k_2, \\ 0 & \text{otherwise.} \end{cases} \end{gathered}$$

The Fourier–Whittaker expansions of $\varphi_{j,k}$, f_k, and E_{2k} are given in Section 4A.

3C. ***Spectral decomposition.*** The spectral decomposition of $L^2(\mathrm{SL}_2(\mathbb{Z})\backslash \mathrm{SL}_2(\mathbb{R}))$ is stated below, which follows by combining [Bump 1997, Corollary to Theorem 2.3.4] with [Duke et al. 2002, Propositions 4.1, 4.2, and 4.3]; for a general reference in the adèlic setting, see [Wu 2017, Theorem 1.3]. Given $a \in$

$L^2(\mathrm{SL}_2(\mathbb{Z})\backslash \mathrm{SL}_2(\mathbb{R}))$, we have the spectral decomposition

$$\begin{aligned} a(z,\theta) &= \frac{3}{\pi}\langle a, 1\rangle + \sum_{\ell=1}^{\infty}\sum_{k=-\infty}^{\infty}\langle a, \Phi_{\ell,k}\rangle \Phi_{\ell,k}(z,\theta) + \sum_{\ell=1}^{\infty}\sum_{F\in\mathcal{H}_\ell}\sum_{\substack{k=-\infty\\ |k|\geq\ell}}^{\infty}\langle a, \Psi_{F,k}\rangle\Psi_{F,k}(z,\theta) \\ &\quad + \frac{1}{4\pi}\sum_{k=-\infty}^{\infty}\int_{-\infty}^{\infty}\bigl\langle a, \widetilde{E}_{2k}\bigl(\cdot,\cdot,\tfrac{1}{2}+it\bigr)\bigr\rangle \widetilde{E}_{2k}\bigl(z,\theta,\tfrac{1}{2}+it\bigr)\,\mathrm{d}t. \end{aligned}$$

This converges in the L^2-sense. If moreover a is smooth and compactly supported, then this converges absolutely and uniformly on compact sets.

We additionally have Parseval's identity: for $a_1, a_2 \in L^2(\mathrm{SL}_2(\mathbb{Z})\backslash \mathrm{SL}_2(\mathbb{R}))$, we have the absolutely convergent spectral expansion

$$\begin{aligned} \langle a_1, a_2\rangle &= \frac{3}{\pi}\langle a_1, 1\rangle\langle 1, a_2\rangle + \sum_{\ell=1}^{\infty}\sum_{k=-\infty}^{\infty}\langle a_1, \Phi_{\ell,k}\rangle\langle\Phi_{\ell,k}, a_2\rangle \\ &\quad + \sum_{\ell=1}^{\infty}\sum_{F\in\mathcal{H}_\ell}\sum_{\substack{k=-\infty\\ |k|\geq\ell}}^{\infty}\langle a_1, \Psi_{F,k}\rangle\langle\Psi_{F,k}, a_2\rangle \\ &\quad + \frac{1}{4\pi}\sum_{k=-\infty}^{\infty}\int_{-\infty}^{\infty}\bigl\langle a_1, \widetilde{E}_{2k}\bigl(\cdot,\cdot,\tfrac{1}{2}+it\bigr)\bigr\rangle\bigl\langle\widetilde{E}_{2k}\bigl(\cdot,\cdot,\tfrac{1}{2}+it\bigr), a_2\bigr\rangle\,\mathrm{d}t. \end{aligned} \tag{14}$$

3D. ***QUE on the modular surface.*** There is a significantly simpler formula for the microlocal lift ω_j of φ_j to a measure on $\mathrm{SL}_2(\mathbb{Z})\backslash \mathrm{SL}_2(\mathbb{R})$. We again recall from [Zelditch 1991, (1.18)] that

$$\mathrm{d}\omega_j(z,\theta) := \varphi_j(z)\overline{u_j(z,\theta)}\,\mathrm{d}\omega(z,\theta), \quad u_j(z,\theta) := \frac{3}{\pi}\sum_{k=-\infty}^{\infty}\varphi_{j,k}(z)e^{2ki\theta},$$

where convergence of the sum defining u_j is in distribution (i.e., $\varphi_j\overline{u_j}\,\mathrm{d}\omega$ is the limit of measures of the partial sums defining u_j).[4] In particular, we have that

$$\int_{\mathrm{SL}_2(\mathbb{Z})\backslash \mathrm{SL}_2(\mathbb{R})}\Phi_{\ell,k}(z,\theta)\,\mathrm{d}\omega_j(z,\theta) = \frac{3}{\pi}\int_{\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}}\varphi_j(z)\overline{\varphi_{j,k}(z)}\varphi_{\ell,k}(z)\,\mathrm{d}\mu(z), \tag{15}$$

$$\int_{\mathrm{SL}_2(\mathbb{Z})\backslash \mathrm{SL}_2(\mathbb{R})}\Psi_{F,k}(z,\theta)\,\mathrm{d}\omega_j(z,\theta) = \frac{3}{\pi}\int_{\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}}\varphi_j(z)\overline{\varphi_{j,k}(z)}f_k(z)\,\mathrm{d}\mu(z), \tag{16}$$

[4] More precisely, the measure is defined by $\int a\,\mathrm{d}\omega_j = \lim_{K\to\infty}\frac{3}{\pi}\int a\varphi_j\sum_{k=-K}^{K}\overline{\varphi_{j,k}}e^{2ki\theta}\,\mathrm{d}\omega$.

and

$$\int_{\mathrm{SL}_2(\mathbb{Z})\backslash\,\mathrm{SL}_2(\mathbb{R})} \widetilde{E}_{2k}\big(z, \theta, \tfrac{1}{2} + it\big)\, d\omega_j(z, \theta) = \frac{3}{\pi} \int_{\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}} \varphi_j(z)\overline{\varphi_{j,k}(z)} E_{2k}\big(z, \tfrac{1}{2} + it\big)\, d\mu(z). \tag{17}$$

Using the spectral decomposition for $L^2(\mathrm{SL}_2(\mathbb{Z})\backslash\,\mathrm{SL}_2(\mathbb{R}))$ and (15), (16), and (17), for any $a \in C_{c,K}^{\infty}(\mathrm{SL}_2(\mathbb{Z})\backslash\,\mathrm{SL}_2(\mathbb{R}))$, we may therefore write

$$\begin{aligned}
&\int_{\mathrm{SL}_2(\mathbb{Z})\backslash\,\mathrm{SL}_2(\mathbb{R})} a(z, \theta)\, d\omega_j(z, \theta) \\
&\quad = \int_{\mathrm{SL}_2(\mathbb{Z})\backslash\,\mathrm{SL}_2(\mathbb{R})} a(z, \theta)\, d\omega(z, \theta) \\
&\qquad + \frac{3}{\pi} \sum_{\ell=1}^{\infty} \sum_{k=-\infty}^{\infty} \langle a, \Phi_{\ell,k}\rangle \int_{\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}} \varphi_j(z)\overline{\varphi_{j,k}(z)}\varphi_{\ell,k}(z)\, d\mu(z) \\
&\qquad + \frac{3}{\pi} \sum_{\ell=1}^{\infty} \sum_{F \in \mathcal{H}_\ell} \sum_{\substack{k=-\infty \\ |k| \geq \ell}}^{\infty} \langle a, \Psi_{F,k}\rangle \int_{\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}} \varphi_j(z)\overline{\varphi_{j,k}(z)} f_k(z)\, d\mu(z) \\
&\qquad + \frac{3}{4\pi^2} \sum_{k=-\infty}^{\infty} \int_{-\infty}^{\infty} \big\langle a, \widetilde{E}_{2k}\big(\cdot\,, \cdot\,, \tfrac{1}{2} + it\big)\big\rangle \\
&\qquad\qquad \times \int_{\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}} \varphi_j(z)\overline{\varphi_{j,k}(z)} E_{2k}\big(z, \tfrac{1}{2} + it\big)\, d\mu(z)\, dt.
\end{aligned} \tag{18}$$

To establish Theorem 1.1, it therefore suffices to bound each of the three integrals (15), (16), and (17). The next few sections will be dedicated to resolving each individual case.

4. Relevant tools for computation

4A. ***Fourier–Whittaker expansions.*** We explicitly write out the Fourier–Whittaker expansion for shifted Hecke–Maaß cusp forms, shifted holomorphic or antiholomorphic Hecke cusp forms, and weighted Eisenstein series. These involve Whittaker functions $W_{\alpha,\beta}(y)$, which are certain special functions on $\mathbb{R}_+$ associated to a pair of parameters $\alpha, \beta \in \mathbb{C}$ that decay exponentially as y tends to infinity in the sense that $\lim_{y\to\infty} y^{-\alpha} e^{y/2} W_{\alpha,\beta}(y) = 1$ (see [Whittaker and Watson 1996, Chapter 16] and [Gradshteyn and Ryzhik 2015, Sections 9.22–9.23]); they satisfy the second order linear ordinary differential equation

$$W''_{\alpha,\beta}(y) + \left(-\frac{1}{4} + \frac{\alpha}{y} + \frac{\frac{1}{4} - \beta^2}{y^2}\right) W_{\alpha,\beta}(y) = 0.$$

In particular cases, these Whittaker functions are simpler: for $\alpha \in \mathbb{C}$, we have that $W_{\alpha,\alpha-1/2}(y) = y^{\alpha} e^{-y/2}$, while for $\beta \in \mathbb{C}$, we have that $W_{0,\beta}(y) = \sqrt{y/\pi}\, K_{\beta}(y/2)$, where $K_{\beta}(y)$ denotes the modified Bessel function of the second kind:

• For Hecke–Maaß cusp forms of weight 0, we have by [Goldfeld and Hundley 2011, Theorem 3.11.8] the Fourier expansion

$$\varphi_j(z) = \sum_{\substack{n=-\infty \\ n\neq 0}}^{\infty} \operatorname{sgn}(n)^{\kappa_j} \rho_j(1) \frac{\lambda_j(|n|)}{\sqrt{|n|}} W_{0,ir_j}(4\pi|n|y) e(nx). \tag{19}$$

Here $\lambda_j(n)$ is the n-th Hecke eigenvalue of φ_j, $\kappa_j \in \{0, 1\}$ is as in (12), and the first Fourier coefficient $\rho_j(1) \in \mathbb{R}_+$ satisfies

$$\rho_j(1)^2 = \frac{\cosh \pi r_j}{2L(1, \operatorname{ad}\varphi_j)} = \frac{\pi}{2\Gamma\left(\frac{1}{2}+ir_j\right)\Gamma\left(\frac{1}{2}-ir_j\right)L(1, \operatorname{ad}\varphi_j)}. \tag{20}$$

By the Rankin–Selberg method (see [Duke et al. 2002, Section 19]), this ensures that φ_j is L^2-normalized. The positive constant $L(1, \operatorname{ad}\varphi_j)$ is the value at $s = 1$ of the adjoint L-function $L(s, \operatorname{ad}\varphi_j)$ defined in (28) below. One can use the recurrence relations for Whittaker functions [Gradshteyn and Ryzhik 2015, (9.234)] to establish that for shifted Maaß cusp forms of weight $2k$,

$$\varphi_{j,k}(z) = \sum_{\substack{n=-\infty \\ n\neq 0}}^{\infty} D_{k,r_j}^{\operatorname{sgn}(n)} \operatorname{sgn}(n)^{\kappa_j} \rho_j(1) \frac{\lambda_j(|n|)}{\sqrt{|n|}} W_{\operatorname{sgn}(n)k,ir_j}(4\pi|n|y) e(nx), \tag{21}$$

where we define the constants

$$D_{k,r}^{\pm} := \frac{(-1)^k \Gamma\left(\frac{1}{2}+ir\right)}{\Gamma\left(\frac{1}{2}\pm k+ir\right)} \tag{22}$$

for $r \in \mathbb{C}$ and $k \in \mathbb{Z}$. One sees from [Duke et al. 2002, Corollary 4.4] that $\varphi_{j,k}$ is also L^2-normalized.

• For shifted holomorphic Hecke cusp forms of positive weight $2k$, we may write the unshifted form as $f = y^{\ell} F$ for some holomorphic Hecke cusp form F of weight 2ℓ. Once more by [Goldfeld and Hundley 2011, Theorem 3.11.8], this has the Fourier expansion

$$f(z) = \sum_{n=1}^{\infty} \rho_F(1) \frac{\lambda_F(n)}{\sqrt{n}} (4\pi n y)^{\ell} e(nz) = \sum_{n=1}^{\infty} \rho_F(1) \frac{\lambda_F(n)}{\sqrt{n}} W_{\ell,\ell-1/2}(4\pi n y) e(nx),$$

where again $\lambda_F(n)$ is the n-th Hecke eigenvalue of F and the first Fourier coefficient $\rho_F(1) \in \mathbb{R}_+$ satisfies

$$\rho_F(1)^2 = \frac{\pi}{2\Gamma(2\ell)L(1, \operatorname{ad} F)}, \tag{23}$$

which ensures that f is L^2-normalized by the Rankin–Selberg method. Once more, the positive constant $L(1, \mathrm{ad}\, F)$ is the value at $s = 1$ of the adjoint L-function $L(s, \mathrm{ad}\, F)$ defined in (28) below. Applying raising operators, we have that

$$(R_{2k-2} \cdots R_{2\ell+2} R_{2\ell} f)(z) = (-1)^{k-\ell} \sum_{n=1}^{\infty} \rho_F(1) \frac{\lambda_F(n)}{\sqrt{n}} W_{k,\ell-1/2}(4\pi n y) e(nx).$$

Finally, we see from [Duke et al. 2002, Corollary 4.4 and (4.60)] that in order to L^2-normalize such a form, we have the final Fourier expansion

$$f_k(z) = \sum_{n=1}^{\infty} C_{k,\ell} \rho_F(1) \frac{\lambda_F(n)}{\sqrt{n}} W_{k,\ell-\frac{1}{2}}(4\pi n y) e(nx) \tag{24}$$

with

$$C_{k,\ell} := (-1)^{k-\ell} \sqrt{\frac{\Gamma(2\ell)}{\Gamma(k+\ell)\Gamma(k-\ell+1)}}. \tag{25}$$

Similarly, for shifted antiholomorphic Hecke cusp forms of negative weight $-2k$, we may write the unshifted Hecke cusp form as $\bar{f} = y^\ell \bar{F}$. One has

$$\overline{f(z)} = \sum_{n=1}^{\infty} \rho_F(1) \frac{\lambda_F(n)}{\sqrt{n}} W_{\ell,\ell-1/2}(4\pi n y) e(-nx),$$

so that

$$f_{-k}(z) = \sum_{n=1}^{\infty} C_{k,\ell} \rho_F(1) \frac{\lambda_F(n)}{\sqrt{n}} W_{k,\ell-1/2}(4\pi n y) e(-nx). \tag{26}$$

- Finally we recall the Fourier expansion of Eisenstein series. Define

$$\lambda(n,t) := \sum_{ab=n} a^{it} b^{-it}.$$

For weight 0 Eisenstein series, we have from [Jakobson 1994, (1.3)] that

$$E\big(z, \tfrac{1}{2}+it\big) = y^{1/2+it} + \frac{\xi(1-2it)}{\xi(1+2it)} y^{1/2-it} + \sum_{\substack{n=-\infty \\ n\neq 0}}^{\infty} \frac{1}{\xi(1+2it)} \frac{\lambda(|n|,t)}{\sqrt{|n|}} W_{0,it}(4\pi |n| y) e(nx), \tag{27}$$

where $\xi(s) := \pi^{-s/2}\Gamma\big(\frac{s}{2}\big)\zeta(s)$ is the completed Riemann zeta function. For weight $2k$ Eisenstein series, we then have that

$$E_{2k}\big(z, \tfrac{1}{2}+it\big) = y^{1/2+it} + \frac{(-1)^k\Gamma\big(\frac{1}{2}+it\big)^2}{\Gamma\big(\frac{1}{2}-k+it\big)\Gamma\big(\frac{1}{2}+k+it\big)}\frac{\xi(1-2it)}{\xi(1+2it)}y^{1/2-it} + \sum_{\substack{n=-\infty \\ n\neq 0}}^{\infty} \frac{D_{k,t}^{\operatorname{sgn}(n)}}{\xi(1+2it)}\frac{\lambda(|n|,t)}{\sqrt{|n|}} W_{\operatorname{sgn}(n)k,it}(4\pi|n|y)e(nx).$$

4B. *L-Functions.* We give a quick overview of all the necessary theory surrounding L-functions. A general discussion of the theory of L-functions and their bounds can be found in [Iwaniec and Kowalski 2004, Chapter 5].

Let ϕ be either a Hecke–Maaß cusp form or a holomorphic Hecke cusp form. Such a Hecke cusp form ϕ has an associated L-function $L(s,\phi)$. Since the Hecke operators T_n satisfy the multiplicativity relation

$$T_m T_n = \sum_{d\,|\,(m,n)} T_{mn/d^2},$$

the Hecke eigenvalues $\lambda_\phi(n)$ must satisfy the corresponding Hecke relations

$$\lambda_\phi(m)\lambda_\phi(n) = \sum_{d\,|\,(m,n)} \lambda_\phi\left(\frac{mn}{d^2}\right).$$

We may therefore define for $\Re(s) > 1$ the degree 2 L-function

$$L(s,\phi) := \sum_{n=1}^{\infty} \frac{\lambda_\phi(n)}{n^s} = \prod_p \frac{1}{1-\lambda_\phi(p)p^{-s}+p^{-2s}}.$$

This can be analytically continued to a holomorphic function on $\mathbb{C}$. We may write the Euler product as

$$L(s,\phi) = \prod_p \frac{1}{(1-\alpha_{\phi,1}(p)p^{-s})^{-1}(1-\alpha_{\phi,2}(p)p^{-s})},$$

where the *Satake parameters* $\alpha_{\phi,1}(p), \alpha_{\phi,2}(p)$ satisfy

$$\alpha_{\phi,1}(p)+\alpha_{\phi,2}(p) = \lambda_\phi(p), \quad \alpha_{\phi,1}(p)\alpha_{\phi,2}(p) = 1.$$

We also define relevant higher degree L-functions: for $m \leq 3$, we define the degree 2^m L-function

$$L(s, \phi_1\otimes\cdots\otimes\phi_m) := \prod_p \prod_{(b_j)\in\{1,2\}^m} \frac{1}{1-\alpha_{\phi_1,b_1}(p)\cdots\alpha_{\phi_m,b_m}(p)p^{-s}}.$$

We additionally define the degree 3 and degree 6 L-functions

$$L(s, \operatorname{ad}\phi) := \frac{L(s, \phi \otimes \phi)}{\zeta(s)}, \tag{28}$$

$$L(s, \operatorname{ad}\phi_1 \otimes \phi_2) := \frac{L(s, \phi_1 \otimes \phi_1 \otimes \phi_2)}{L(s, \phi_2)}, \tag{29}$$

Each of these L-functions has a meromorphic continuation to $\mathbb{C}$. For later use, we will also recall the identities

$$\sum_{n=1}^{\infty} \frac{\lambda_\phi(n)^2}{n^s} = \frac{\zeta(s)L(s, \operatorname{ad}\phi)}{\zeta(2s)}, \tag{30}$$

$$\sum_{n=1}^{\infty} \frac{\lambda_\phi(n)\lambda(n,t)}{n^s} = \frac{L(s+it, \phi)L(s-it, \phi)}{\zeta(2s)}, \tag{31}$$

which are both valid for $\Re(s) > 1$.

For any such L-function $L(s, \Pi)$ of degree d, where Π is a placeholder for one of the automorphic objects listed above, we have a corresponding gamma factor of the form

$$L_\infty(s, \Pi) = \prod_{i=1}^{d} \Gamma_{\mathbb{R}}(s + \mu_i)$$

for some *Langlands parameters* $\mu_i \in \mathbb{C}$, where $\Gamma_{\mathbb{R}}(s) := \pi^{-s/2}\Gamma\left(\frac{s}{2}\right)$. The completed L-function $\Lambda(s, \Pi) := L(s, \Pi)L_\infty(s, \Pi)$ has a meromorphic continuation to $\mathbb{C}$ and satisfies a functional equation of the form $\Lambda(1-s, \Pi) = \epsilon_\Pi \Lambda(s, \widetilde{\Pi})$, where the *epsilon factor* ϵ_Π is a complex number of absolute value 1, while $\Lambda(s, \widetilde{\Pi}) = \overline{\Lambda(\bar{s}, \Pi)}$.

4C. ***Bounds for L-Functions.*** Various L-functions will appear in the integrals computed later in the paper. As such, the study of the sizes of our integrals is connected to the study of the sizes of such L-functions. In particular, estimating relevant integrals can be reduced to estimating $L(1, \Pi)$ and $L\left(\frac{1}{2}+it, \Pi\right)$ for various values of t and Π. We discuss the specific relevant bounds.

For ϕ a Hecke–Maaß cusp form with spectral parameter r, combining the work of [Goldfeld et al. 1994, main theorem] and [Li 2010, Corollary 1] with (20), we have that

$$\frac{1}{\log r} \ll L(1, \operatorname{ad}\phi) \ll \exp(C(\log r)^{1/4}(\log\log r)^{1/2}) \tag{32}$$

for some absolute constant $C > 0$. Similarly, for ϕ a holomorphic Hecke cusp form of weight ℓ, we have that

$$\frac{1}{\log \ell} \ll L(1, \operatorname{ad}\phi) \ll (\log \ell)^3, \tag{33}$$

where the lower bound again follows from [Goldfeld et al. 1994, main theorem], while the upper bound follows from [Lau and Wu 2006, Proposition 3.2(i)]. Finally, for $t \in \mathbb{R}$, we have the classical bounds [Iwaniec and Kowalski 2004, (8.24), Theorem 8.29]

$$\frac{1}{(\log(3+|t|))^{2/3}(\log\log(9+|t|))^{1/3}} \ll |\zeta(1+it)| \ll \frac{\log(3+|t|)}{\log\log(9+|t|)}. \tag{34}$$

We shall only make use of the lower bounds in (32), (33), and (34). In particular, the lower bound in (32) is precisely the cause of the presence of the term $\log \lambda_j$ on the right-hand side of (6).

To discuss values of an L-function $L(s, \Pi)$ on the line $\Re(s) = \frac{1}{2}$, we define the *analytic conductor*

$$C(s, \Pi) := \prod_{i=1}^{d}(1+|s+\mu_i|).$$

The analytic conductor can be thought of as measuring the *complexity* of the L-function $L(s, \Pi)$. The *convexity bound* bound for such an L-function on the line $\Re(s) = \frac{1}{2}$ is

$$L(s, \Pi) \ll_\varepsilon C(s, \Pi)^{1/4+\varepsilon}.$$

Such a bound is known for all of the L-functions that we study below; it is a consequence of the Phragmén–Lindelöf convexity principle, the functional equations for these L-functions, and upper bounds for these L-functions at the edge of the critical strip [Li 2010, Theorem 2]. A *subconvex* bound is a bound of the form

$$L(s, \Pi) \ll C(s, \Pi)^{1/4-\delta}$$

for some fixed $\delta > 0$; in contrast, for some of the L-functions that we study below, such a bound is not yet known. The *generalized Lindelöf hypothesis* is the conjecture that such a subconvex bound holds with $\delta = \frac{1}{4} - \varepsilon$ for any fixed $\varepsilon > 0$. The generalized Lindelöf hypothesis would follow as a consequence from the *generalized Riemann hypothesis*, which is the conjecture that the only zeroes of $L(s, \Pi)$ in the critical strip $0 < \Re(s) < 1$ lie on the critical line $\Re(s) = \frac{1}{2}$.

We make this explicit for various L-functions of interest to us by recalling the values of the Langlands parameters μ_i in these cases. An elementary example is the Riemann zeta function, which is of degree 1: the Langlands parameter is simply $\mu_1 = 0$, so that the convexity bound is

$$\zeta\left(\tfrac{1}{2}+it\right) \ll_\varepsilon (1+|t|)^{1/4+\varepsilon}. \tag{35}$$

Next, from [Iwaniec and Kowalski 2004, Sections 5.11 and 5.12], when φ and $\widetilde{\varphi}$ are Maaß cusp forms with spectral parameters r and $\tilde{r}$ and parities ϵ and $\widetilde{\epsilon}$, we have that

$$\begin{aligned} L_\infty(s,\varphi) &= \Gamma_{\mathbb{R}}\Big(s+\frac{1-\epsilon}{2}+ir\Big)\Gamma_{\mathbb{R}}\Big(s+\frac{1-\epsilon}{2}-ir\Big), \\ L_\infty(s,\operatorname{ad}\varphi) &= \Gamma_{\mathbb{R}}(s+2ir)\Gamma_{\mathbb{R}}(s)\Gamma_{\mathbb{R}}(s-2ir) \\ L_\infty(s,\operatorname{ad}\varphi\otimes\widetilde{\varphi}) &= \prod_{\pm}\Gamma_{\mathbb{R}}\Big(s+\frac{1-\widetilde{\epsilon}}{2}+2ir\pm i\tilde{r}\Big)\Gamma_{\mathbb{R}}\Big(s+\frac{1-\widetilde{\epsilon}}{2}\pm i\tilde{r}\Big) \\ &\qquad\qquad \times\Gamma_{\mathbb{R}}\Big(s+\frac{1-\widetilde{\epsilon}}{2}-2ir\pm i\tilde{r}\Big). \end{aligned}$$

In particular, we have the convexity bounds

$$L\big(\tfrac{1}{2},\varphi\big)\ll_\varepsilon r^{1/2+\varepsilon}, \tag{36}$$

$$L\big(\tfrac{1}{2}+it,\operatorname{ad}\varphi\big)\ll_\varepsilon ((1+|t|)(1+|r+t|)(1+|r-t|))^{1/4+\varepsilon}, \tag{37}$$

$$L\big(\tfrac{1}{2},\operatorname{ad}\varphi\otimes\widetilde{\varphi}\big)\ll_\varepsilon (\tilde{r}(r+\tilde{r})(1+|r-\tilde{r}|))^{1/2+\varepsilon}. \tag{38}$$

For our applications regarding QUE, we need to assume hypothetical improvements upon (37) and (38) that imply subconvexity in the r-aspect but allow for polynomial growth in the t-aspect or $\tilde{r}$-aspect, namely bounds of the form

$$\begin{aligned} L\big(\tfrac{1}{2}+it,\operatorname{ad}\varphi\big) &\ll r^{1/2-2\delta}(1+|t|)^A, \\ L\big(\tfrac{1}{2},\operatorname{ad}\varphi\otimes\widetilde{\varphi}\big) &\ll r^{1-4\delta}\tilde{r}^{2A} \end{aligned}$$

for some $\delta>0$ and $A>0$ (see Theorems 6.2 and 8.2).

Finally, when φ is again a Maaß cusp form with spectral parameter r and F is a holomorphic Hecke cusp form of weight $2\ell>0$, we have that

$$\begin{aligned} L_\infty(s,F) &= \Gamma_{\mathbb{R}}\big(s+\ell+\tfrac{1}{2}\big)\Gamma_{\mathbb{R}}\big(s+\ell-\tfrac{1}{2}\big) \\ L_\infty(s,\operatorname{ad}\varphi\otimes F) &= \prod_{\pm}\Gamma_{\mathbb{R}}\big(s+2ir+\ell\pm\tfrac{1}{2}\big)\Gamma_{\mathbb{R}}\big(s+\ell\pm\tfrac{1}{2}\big)\Gamma_{\mathbb{R}}\big(s-2ir+\ell\pm\tfrac{1}{2}\big). \end{aligned}$$

In particular, we have the convexity bounds

$$L\big(\tfrac{1}{2}+it,F\big)\ll_\varepsilon (\ell+|t|)^{1/2+\varepsilon}, \tag{39}$$

$$L\big(\tfrac{1}{2},\operatorname{ad}\varphi\otimes F\big)\ll_\varepsilon (\ell(r+\ell)^2)^{1/2+\varepsilon}. \tag{40}$$

Good [1982, Corollary] has proven an improvement upon (39) that implies subconvexity in the t-aspect, namely the subconvex bound

$$L\big(\tfrac{1}{2}+it,F\big)\ll_{\ell,\varepsilon} |t|^{1/3+\varepsilon}. \tag{41}$$

For our applications regarding QUE, we also need to assume a hypothetical improvement upon (40) that implies subconvexity in the r-aspect, namely a bound of the form

$$L\left(\tfrac{1}{2}, \operatorname{ad}\varphi \otimes F\right) \ll_\ell r^{1-4\delta}$$

for some $\delta > 0$ (see Theorem 9.2).

5. Completing the proof of continuous spectrum QUE

We now supply the necessary computation missing from Jakobson's proof of QUE for Eisenstein series given in [Jakobson 1994]. This setting shares many similarities with that of QUE for Hecke–Maaß cusp forms; the chief alteration is that the microlocal lift

$$\omega_j(z,\theta) := \frac{3}{\pi}\varphi_j(z) \sum_{k=-\infty}^{\infty} \overline{\varphi_{j,k}(z) e^{2ki\theta}}$$

of a Hecke–Maaß cusp form φ_j is replaced by the microlocal lift

$$\mu_t(z,\theta) := \frac{3}{\pi} E\left(z, \tfrac{1}{2}+it\right) \sum_{k=-\infty}^{\infty} \overline{E_{2k}\left(z, \tfrac{1}{2}+it\right) e^{2ki\theta}}$$

of an Eisenstein series $E\left(z, \frac{1}{2}+it\right)$. Similar to the discussion in Section 3D, Jakobson's proof of QUE for Eisenstein series requires one to bound both of the integrals

$$\int_{\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}} E\left(z, \tfrac{1}{2}+it\right) E_{-2k}\left(z, \tfrac{1}{2}-it\right) \varphi_{\ell,k}(z)\, \mathrm{d}\mu(z), \tag{42}$$

where $\varphi_{\ell,k}$ is a shifted Hecke–Maaß cusp form of weight $2k \geq 0$ arising from a Hecke–Maaß cusp form φ_ℓ of weight 0 and spectral parameter r_ℓ, and

$$\int_{\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}} E\left(z, \tfrac{1}{2}+it\right) E_{-2k}\left(z, \tfrac{1}{2}-it\right) f_k(z)\, \mathrm{d}\mu(z), \tag{43}$$

where f_k is a shifted holomorphic Hecke cusp form of weight $2k > 0$ obtained by raising a holomorphic Hecke cusp form F of weight $2\ell > 0$ with $\ell < k$. One must similarly also bound an integral involving three Eisenstein series, of which two are shifted; this requires some minor alterations involving incomplete Eisenstein series (see [Jakobson 1994, Section 3]), since otherwise this integral would diverge.

Jakobson treats this altered Eisenstein integral in [loc. cit., Proposition 3.1], while he treats the shifted Hecke–Maaß cusp form integral (42) in [loc. cit., Proposition 2.2]. For the shifted holomorphic Hecke cusp form integral (43), Jakobson only treats the *unshifted* case in [loc. cit., Proposition 2.1].

To treat the shifted case, we first relate an integral of two Eisenstein series and a shifted holomorphic Hecke cusp form to the product of a ratio of L-functions and an integral involving Whittaker functions.

Lemma 5.1. *For any shifted holomorphic Hecke cusp form f_k of weight $2k > 0$ obtained by raising a holomorphic Hecke cusp form F of weight $2\ell > 0$ with $\ell < k$, we have that*

$$\begin{aligned}\int_{\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}} & E\big(z, \tfrac{1}{2}+it\big)E_{-2k}\big(z, \tfrac{1}{2}-it\big) f_k(z)\,\mathrm{d}\mu(z) \\ &= (-1)^{k-\ell}\sqrt{\frac{\pi}{2}}(2\pi)^{1+2it}\frac{L\big(\frac{1}{2}, F\big)L\big(\frac{1}{2}-2it, F\big)}{\zeta(1-2it)\zeta(1+2it)\sqrt{L(1, \operatorname{ad} F)}} \\ &\qquad \times \int_0^\infty \frac{W_{0,it}(u)}{\Gamma\big(\frac{1}{2}+it\big)}\frac{W_{k,\ell-1/2}(u)}{\sqrt{\Gamma(k+\ell)\Gamma(k-\ell+1)}}u^{-1/2-it}\,\frac{\mathrm{d}u}{u}. \end{aligned} \tag{44}$$

Proof. We begin by studying the integral

$$I_1(s) := \int_{\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}} E\big(z, \tfrac{1}{2}+it\big)E_{-2k}(z, s) f_k(z)\,\mathrm{d}\mu(z)$$

when $\Re(s) > 1$, which allows use to make use of the absolutely convergent expression (13) for $E_{-2k}(z, s)$; we later analytically continue $I_1(s)$ to $s = \frac{1}{2} - it$. We first apply the unfolding trick, inserting the identity (13) for $E_{-2k}(z, s)$ and turning the integral over $\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}$ into one over $\Gamma_\infty\backslash\mathbb{H}$; see [Jakobson 1994, Proof of Proposition 2.1]. Using the fact that f_k has weight $2k$, we have that

$$I_1(s) = \int_{\Gamma_\infty\backslash\mathbb{H}} E\big(z, \tfrac{1}{2}+it\big) f_k(z)\Im(z)^s\,\mathrm{d}\mu(z).$$

We evaluate this integral by taking a fundamental domain of $\Gamma_\infty\backslash\mathbb{H}$ to be $[0, 1]\times\mathbb{R}_+$. We now insert the Fourier–Whittaker expansions (27) of $E\big(z, \frac{1}{2}+it\big)$ and (24) of $f_k(z)$, interchange the order of summation and integration, evaluate the integral over $x \in [0, 1]$, and make the substitution $u = 4\pi|n|y$. This leads us to the identity

$$I_1(s) = \frac{(4\pi)^{1-s}C_{k,\ell}\rho_F(1)}{\xi(1+2it)}\sum_{n=1}^{\infty}\frac{\lambda_F(n)\lambda(n, t)}{n^s}\int_0^\infty W_{0,it}(u)W_{k,\ell-1/2}(u)u^{s-1}\,\frac{\mathrm{d}u}{u}.$$

At this point, we analytically continue this expression to $s = \frac{1}{2} - it$, as the Dirichlet series extends holomorphically to the closed half-plane $\Re(s) \geq \frac{1}{2}$ from (31) (recalling that $\zeta(2s) \neq 0$ for $\Re(s) \geq \frac{1}{2}$), while the integral extends holomorphically to the open half-plane $\Re(s) > \frac{1}{2} - \ell$ by [Gradshteyn and Ryzhik 2015, (7.621.11) and (9.237.3)], since these identities allow us to write the integral as a finite sum of quotients of gamma functions that have no poles for $\Re(s) > \frac{1}{2} - \ell$. Recalling the identities (25) for $C_{k,\ell}$, (23) for $\rho_F(1)^2$, and (31) for the Dirichlet series, we obtain the desired identity. □

Theorem 5.2. *For any shifted holomorphic or antiholomorphic Hecke cusp form f_k of weight $2k$ obtained by raising or lowering a holomorphic Hecke cusp form F of weight $2\ell > 0$ with $\ell < |k|$, we have that*

$$\int_{\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}} E\bigl(z, \tfrac{1}{2}+it\bigr)E_{-2k}\bigl(z, \tfrac{1}{2}-it\bigr) f_k(z)\,\mathrm{d}\mu(z) \ll_{k,\ell,\varepsilon} |t|^{-1/6+\varepsilon}.$$

Proof. We consider only the positive weight case; the analogous bounds for the negative weight case follow by conjugational symmetry. We bound the expression (44). Via (34) and the subconvex bound (41), the ratio of L-functions is $O_{\ell,\varepsilon}(|t|^{1/3+\varepsilon})$. It remains to deal with the integral of Whittaker functions. In Corollary A.7, we show that this integral is $O_{k,\ell}(|t|^{-1/2})$. This yields the desired estimate. □

Remark 5.3. Theorem 5.2 is unconditional due to the fact that the subconvex bound (41) for $L\bigl(\tfrac{1}{2}+it, F\bigr)$ is known unconditionally. A similar such subconvex bound is known for $L\bigl(\tfrac{1}{2}+it, \varphi\bigr)$, where φ is a *fixed* Hecke–Maaß cusp form [Meurman 1990]; finally, an analogous subconvex bound is known for $\zeta\bigl(\tfrac{1}{2}+it\bigr)$. These known subconvex bounds are the key inputs for Jakobson's unconditional proof of QUE for Eisenstein series [Jakobson 1994, Theorem 1]. In contrast, the subconvex bounds (3), (4), and (5) remain hypothetical, which is the reason that Theorem 1.1 is a conditional result.

6. Eisenstein series computation

We now move on to the proof of our main theorem, first proving the desired bound for Eisenstein series. We begin by relating an integral of a Hecke–Maaß cusp form, a shifted Hecke–Maaß cusp form, and a shifted Eisenstein series to the product of a ratio of L-functions and an integral involving Whittaker functions.

Lemma 6.1. *For $k \in \mathbb{Z}$ and $t \in \mathbb{R}$, we have that*

$$\begin{aligned}\int_{\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}} &\varphi_j(z)\overline{\varphi_{j,k}(z)}E_{2k}\bigl(z, \tfrac{1}{2}+it\bigr)\,\mathrm{d}\mu(z)\\ &= \frac{\pi}{2}(-1)^{k+\kappa_j}(4\pi)^{1/2-it}\frac{\zeta\bigl(\frac{1}{2}+it\bigr)L\bigl(\frac{1}{2}+it, \mathrm{ad}\,\varphi_j\bigr)}{\zeta(1+2it)L(1, \mathrm{ad}\,\varphi_j)}\\ &\qquad\times\int_0^\infty \frac{W_{0,ir_j}(u)}{\Gamma\bigl(\frac{1}{2}+ir_j\bigr)}\left(\frac{W_{k,-ir_j}(u)}{\Gamma\bigl(\frac{1}{2}+k-ir_j\bigr)}+\frac{W_{-k,-ir_j}(u)}{\Gamma\bigl(\frac{1}{2}-k-ir_j\bigr)}\right)u^{-1/2+it}\,\frac{\mathrm{d}u}{u}.\end{aligned} \tag{45}$$

Proof. We follow the same method as in Lemma 5.1, first evaluating the integral

$$I_2(s) := \int_{\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}} \varphi_j(z)\overline{\varphi_{j,k}(z)}E_{2k}(z, s)\,\mathrm{d}\mu(z)$$

for $\Re(s) > 1$, and then analytically continuing this expression to $s = \frac{1}{2} + it$. We again apply the unfolding trick by inserting the identity (13) for $E_{2k}(z, s)$, giving

$$I_2(s) = \int_{\Gamma_\infty \backslash \mathbb{H}} \varphi_j(z) \overline{\varphi_{j,k}(z)} \Im(z)^s \, d\mu(z).$$

Inserting the Fourier–Whittaker expansions (19) for φ_j and (21) for $\varphi_{j,k}$ and integrating over the fundamental domain $[0, 1] \times \mathbb{R}_+$ of $\Gamma_\infty \backslash \mathbb{H}$, we find that $I_2(s)$ is equal to

$$(-1)^{\kappa_j} (4\pi)^{1-s} \rho_j(1)^2 \\ \times \sum_{n=1}^{\infty} \frac{\lambda_j(n)^2}{n^s} \int_0^\infty W_{0,ir_j}(u) \big(\overline{D^+_{k,-r_j}} W_{k,-ir_j}(u) + \overline{D^-_{k,-r_j}} W_{-k,-ir_j}(u) \big) u^{s-1} \frac{du}{u}.$$

We then analytically continue this to $s = \frac{1}{2} + it$, as the Dirichlet series extends meromorphically to the closed half-plane $\Re(s) \geq \frac{1}{2}$ with only a simple pole at $s = 1$ from (30) (recalling that $\zeta(2s) \neq 0$ for $\Re(s) \geq \frac{1}{2}$), while the integral extends holomorphically to the open half-plane $\Re(s) > 0$ by [Gradshteyn and Ryzhik 2015, (7.611.7)]. Recalling the identities (22) for $D^{\pm}_{k,-r_j}$ (and noting that $\overline{\Gamma(z)} = \Gamma(\bar{z})$), (20) for $\rho_j(1)^2$, and (30) for the Dirichlet series, we obtain the desired identity. □

Theorem 6.2. *For any $\delta > 0$ and $A > 0$, given a subconvex bound of the form*

$$L\big(\tfrac{1}{2} + it, \operatorname{ad} \phi\big) \ll r^{1/2 - 2\delta} (1 + |t|)^A, \tag{46}$$

where ϕ is an arbitrary Hecke–Maaß cusp form with spectral parameter r, we have that

$$\int_{\mathrm{SL}_2(\mathbb{Z}) \backslash \mathbb{H}} \varphi_j(z) \overline{\varphi_{j,k}(z)} E_{2k}\big(z, \tfrac{1}{2} + it\big) \, d\mu(z) \\ \ll_{k,\varepsilon} r_j^{1/2 - 2\delta} \log r_j \, (1 + |t|)^{A - 1/4 + \varepsilon} (2r_j + |t|)^{-1/4} \big(1 + \big|2r_j - |t|\big|\big)^{-1/4}.$$

Proof. We consider only the positive weight case; the analogous bounds for the negative weight case follow by conjugational symmetry. We bound the expression (45). Via the assumption of the subconvex bound (46), the bounds (32) and (34), and the convexity bound (35), the ratio of L-functions in (45) is $O_\varepsilon(r_j^{1/2 - 2\delta} (\log r_j)(1 + |t|)^{A + 1/4 + \varepsilon})$. It remains to deal with the integral of Whittaker functions. In Corollary A.7, we show that this integral is $O_k((1 + |t|)^{-1/2} (2r_j + |t|)^{-1/4} (1 + |2r_j - |t||)^{-1/4})$. This yields the desired estimate. □

7. The Watson–Ichino triple product formula

The remaining integrals we wish to compute are of the form

$$\int_{\mathrm{SL}_2(\mathbb{Z}) \backslash \mathbb{H}} \phi_1(z) \phi_2(z) \phi_3(z) \, d\mu(z)$$

where ϕ_i are (shifted Maaß, holomorphic, or antiholomorphic) Hecke cusp forms of weight $2k_i$ for which $k_1+k_2+k_3=0$. We will compute these via the *Watson–Ichino triple product formula*, which allows us to express these in terms of products of L-functions and integrals of Whittaker functions.

The formula given by Ichino [2008, Theorem 1] is extremely general and simplifies greatly when applied to the special case of cusp forms on the modular surface. We follow the simplification of the general formula done in [Sarnak and Zhao 2019, Appendix].

Let $\widetilde{\phi_i}$ denote the adèlic lift of ϕ_i to a function on $\mathrm{Z}(\mathbb{A}_{\mathbb{Q}})\,\mathrm{GL}_2(\mathbb{Q})\backslash\mathrm{GL}_2(\mathbb{A}_{\mathbb{Q}})$, as described in [Humphries and Nordentoft 2022, Section 4.3]; see also [Goldfeld and Hundley 2011, Section 4.12]. We have that

$$\begin{aligned}\int_{\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}} &\phi_1(z)\phi_2(z)\phi_3(z)\,\mathrm{d}\mu(z)\\ &= \int_{\mathrm{SL}_2(\mathbb{Z})\backslash\mathrm{SL}_2(\mathbb{R})} \phi_1(z)e^{2k_1 i\theta}\phi_2(z)e^{2k_2 i\theta}\phi_3(z)e^{2k_3 i\theta}\,\mathrm{d}\omega(z,\theta)\\ &= \frac{\pi}{6}\int_{\mathrm{Z}(\mathbb{A}_{\mathbb{Q}})\,\mathrm{GL}_2(\mathbb{Q})\backslash\mathrm{GL}_2(\mathbb{A}_{\mathbb{Q}})} \widetilde{\phi_1}(g)\widetilde{\phi_2}(g)\widetilde{\phi_3}(g)\,\mathrm{d}g. \end{aligned}\tag{47}$$

Here $\mathrm{d}g$ denotes the Tamagawa measure on $\mathrm{Z}(\mathbb{A}_{\mathbb{Q}})\,\mathrm{GL}_2(\mathbb{Q})\backslash\mathrm{GL}_2(\mathbb{A}_{\mathbb{Q}})$, which is normalized such that this quotient space has volume 2. The factor $\frac{\pi}{6}$ occurs on the right-hand side of (47) to ensure that the relevant measures are normalized consistently, which can be checked by replacing the integrands with the constant function 1.

The Watson–Ichino triple product formula relates the integral (47) to L-functions and to an integral of matrix coefficients of the local representations of $\mathrm{GL}_2(\mathbb{R})$ associated to $\tilde{\phi}_1, \tilde{\phi}_2, \tilde{\phi}_3$. So long as one of ϕ_1, ϕ_2, ϕ_3 is a shifted Hecke–Maaß cusp form, this integral of matrix coefficients can in turned be expressed as in terms of an integral of *local Whittaker functions* and an element of the *induced model*, which we describe below. The reduction to an integral of this form is a local analogue of the unfolding method used in Lemmas 5.1 and 6.1 and leads to integrals of Whittaker functions of the same form as those appearing in (44) and (45).

7A. ***The Whittaker model.*** Associated to a shifted Maaß, holomorphic, or antiholomorphic Hecke cusp form ϕ of weight $2k$ is a weight $2k$ local Whittaker function $W_\phi : \mathrm{GL}_2(\mathbb{R}) \to \mathbb{C}$. This function satisfies

$$W_\phi\left(\begin{pmatrix}1 & x\\ 0 & 1\end{pmatrix}\begin{pmatrix}y & 0\\ 0 & 1\end{pmatrix}\begin{pmatrix}z & 0\\ 0 & z\end{pmatrix}\begin{pmatrix}\cos\theta & \sin\theta\\ -\sin\theta & \cos\theta\end{pmatrix}\right) = e(x)e^{2ki\theta}W_\phi\begin{pmatrix}y & 0\\ 0 & 1\end{pmatrix} \tag{48}$$

for all $x \in \mathbb{R}$, $y, z \in \mathbb{R}^\times$, and $\theta \in \mathbb{R}$; additionally, letting $\lambda_\phi(n)$ denote the n-th Hecke eigenvalue of ϕ, we have that for $x \in \mathbb{R}$ and $y \in \mathbb{R}_+$,

$$\phi(x+iy) = \sum_{\substack{n=-\infty \\ n\neq 0}}^{\infty} \frac{\lambda_\phi(|n|)}{\sqrt{|n|}} W_\phi \begin{pmatrix} ny & 0 \\ 0 & 1 \end{pmatrix} e(nx); \tag{49}$$

see [Humphries and Nordentoft 2022, Section 4.3.3]. By (21), (24), and (26), this means that $W_\phi\left(\begin{smallmatrix} y & 0 \\ 0 & 1 \end{smallmatrix}\right)$ can be expressed in terms of a constant multiple of a classical Whittaker function $W_{\alpha,\beta}$.

This local Whittaker function W_ϕ is an element of the *Whittaker model* $\mathcal{W}(\pi_\infty)$ associated to ϕ. As explained in [Goldfeld and Hundley 2011, Section 4.8], associated to ϕ is an adèlic automorphic form $\widetilde{\phi} : \mathrm{GL}_2(\mathbb{A}_\mathbb{Q}) \to \mathbb{C}$. In turn, such an adèlic automorphic form is associated to a cuspidal automorphic representation π of $\mathrm{GL}_2(\mathbb{A}_\mathbb{Q})$, as discussed in [loc. cit., Section 5.4]. From [loc. cit., Section 10.4], this automorphic representation is isomorphic to a restricted tensor product of local representations: we have that $\pi \cong \pi_\infty \otimes \bigotimes_p' \pi_p$, where π_∞ is an irreducible representation of $\mathrm{GL}_2(\mathbb{R})$, while π_p is an irreducible representation of $\mathrm{GL}_2(\mathbb{Q}_p)$ for each prime p.

The irreducible representation π_∞ of $\mathrm{GL}_2(\mathbb{R})$ is completed determined by ϕ as follows:

- If ϕ is a shifted Hecke–Maaß cusp form $\varphi_{j,k}$, then π_∞ is a principal series representation. A model for this representation is the Whittaker model $\mathcal{W}(\pi_\infty)$, which consists of certain local Whittaker functions $W : \mathrm{GL}_2(\mathbb{R}) \to \mathbb{C}$, with the irreducible representation π_∞ given via the action $(\pi_\infty(h) \cdot W)(g) := W(gh)$ of $h \in \mathrm{GL}_2(\mathbb{R})$. Letting $r_j \in \mathbb{R}$ denote the spectral parameter of ϕ and $\epsilon_j = (-1)^{\kappa_j}$ denote the parity of ϕ, where $\kappa_j \in \{0, 1\}$, the Whittaker model $\mathcal{W}(\pi_\infty)$ is the vector space of local Whittaker functions $W : \mathrm{GL}_2(\mathbb{R}) \to \mathbb{C}$ of the form

$$W(g) := \mathrm{sgn}(\det g)^{\kappa_j} |\det g|^{1/2+ir_j} \int_{\mathbb{R}^\times} |a|^{-2ir_j} \int_{\mathbb{R}} \Phi((a^{-1}, x)g) e(-ax) \, \mathrm{d}x \, \mathrm{d}^\times a \tag{50}$$

with $\Phi : \mathbb{R}^2 \to \mathbb{C}$ a Schwartz function [Jacquet and Langlands 1970, Lemma 2.5.13.1].

By taking the Schwartz function to be

$$\Phi(x_1, x_2) := \pi^{|k|} \frac{\Gamma\left(\frac{1}{2} + ir_j\right)}{\Gamma\left(\frac{1}{2} + |k| + ir_j\right)} \rho_j(1) (x_2 - \mathrm{sgn}(k) i x_1)^{2|k|} e^{-\pi(x_1^2 + x_2^2)}, \tag{51}$$

the resulting local Whittaker function given by (50) is W_ϕ; it satisfies (48) and is such that

$$W_\phi \begin{pmatrix} y & 0 \\ 0 & 1 \end{pmatrix} = D_{k,r_j}^{\mathrm{sgn}(y)} \, \mathrm{sgn}(y)^{\kappa_j} \rho_j(1) W_{\mathrm{sgn}(y)k, ir_j}(4\pi|y|). \tag{52}$$

This can be seen directly by taking $g = \left(\begin{smallmatrix} y & 0 \\ 0 & 1 \end{smallmatrix}\right)$ in (50) and making the change of variables $x \mapsto a^{-1}xy$ and $a \mapsto \pi^{1/2}|a|^{-1/2}(x^2+1)^{1/2}|y|$, which shows that

$$\begin{aligned} W_\phi \begin{pmatrix} y & 0 \\ 0 & 1 \end{pmatrix} &= \pi^{-1/2-ir_j} \frac{(-1)^k \Gamma\left(\frac{1}{2}+ir_j\right)}{\Gamma\left(\frac{1}{2}+|k|+ir_j\right)} \operatorname{sgn}(y)^{\kappa_j} \rho_j(1)|y|^{1/2-ir_j} \int_0^\infty a^{1/2+|k|+ir_j} e^{-a} \frac{\mathrm{d}a}{a} \\ &\qquad \times \int_{\mathbb{R}} (1+ix)^{-1/2+k-ir_j}(1-ix)^{-1/2-k-ir_j} e(-xy)\,\mathrm{d}x. \end{aligned}$$

The integral over $\mathbb{R}_+ \ni a$ is $\Gamma\left(\frac{1}{2}+|k|+ir_j\right)$, while from [Gradshteyn and Ryzhik 2015, 3.384.9],

$$\begin{aligned} \int_{\mathbb{R}} (1+ix)^{-1/2+k-ir_j}(1-ix)^{-1/2-k-ir_j} e(-xy)\,\mathrm{d}x \\ = \frac{\pi^{1/2+ir_j}|y|^{-1/2+ir_j}}{\Gamma\left(\frac{1}{2}+\operatorname{sgn}(y)k+ir_j\right)} W_{\operatorname{sgn}(y)k,ir_j}(4\pi|y|). \end{aligned}$$

By the definition (22) of $D^{\pm}_{k,r}$, this yields (52).

- If ϕ is a shifted holomorphic or antiholomorphic Hecke cusp form f_k, then π_∞ is a discrete series representation. A model for this representation is the Whittaker model $\mathcal{W}(\pi_\infty)$, which consists of certain local Whittaker functions $W : \mathrm{GL}_2(\mathbb{R}) \to \mathbb{C}$, with the irreducible representation π_∞ given via the action $(\pi_\infty(h) \cdot W)(g) := W(gh)$ of $h \in \mathrm{GL}_2(\mathbb{R})$. Letting $2\ell \in 2\mathbb{N}$ denote the weight of the underlying holomorphic Hecke cusp form F, the Whittaker model $\mathcal{W}(\pi_\infty)$ of π_∞ is the vector space of local Whittaker functions $W : \mathrm{GL}_2(\mathbb{R}) \to \mathbb{C}$ of the form

$$W(g) := |\det g|^{\ell} \int_{\mathbb{R}^\times} |y|^{1-2\ell} \int_{\mathbb{R}} \Phi((y^{-1}, x)g) e(-xy)\,\mathrm{d}x\,\mathrm{d}^\times y \tag{53}$$

with $\Phi : \mathbb{R}^2 \to \mathbb{C}$ a Schwartz function satisfying

$$\int_{\mathbb{R}} x_1^m \left(\frac{\partial^{2\ell-m-2}}{\partial x_2^{2\ell-m-2}} \Big|_{x_2=0} \int_{\mathbb{R}} \Phi(x_1,\xi_2) e(-x_2\xi_2)\,\mathrm{d}\xi_2 \right) \mathrm{d}x_1 = 0$$

for all $m \in \{0, \dots, 2\ell-2\}$ [Jacquet and Langlands 1970, Corollary 2.5.14].

By taking

$$\Phi(x_1,x_2) := \pi^{|k|}(-1)^k C_{|k|,\ell}\,\rho_F(1)(x_2 - \operatorname{sgn}(k) i x_1)^{2|k|} e^{-\pi(x_1^2+x_2^2)},$$

the resulting local Whittaker function given by (53) is W_ϕ; it satisfies (48) and is such that

$$W_\phi \begin{pmatrix} y & 0 \\ 0 & 1 \end{pmatrix} = \begin{cases} C_{|k|,\ell}\,\rho_F(1) W_{|k|,\ell-\frac{1}{2}}(4\pi|y|) & \text{if } \operatorname{sgn}(y) = \operatorname{sgn}(k), \\ 0 & \text{if } \operatorname{sgn}(y) = -\operatorname{sgn}(k). \end{cases} \tag{54}$$

This can again be seen directly by taking $g = \left(\begin{smallmatrix} y & 0 \\ 0 & 1 \end{smallmatrix}\right)$ in (53) and making the change of variables $x \mapsto a^{-1}xy$ and $a \mapsto \pi^{1/2}|a|^{-1/2}(x^2+1)^{1/2}|y|$, which shows that

$$W_\phi \begin{pmatrix} y & 0 \\ 0 & 1 \end{pmatrix} = \pi^{-\ell} C_{|k|,\ell}\rho_F(1)|y|^{1-\ell} \int_0^\infty a^{|k|+\ell} e^{-a} \frac{\mathrm{d}a}{a} \int_{\mathbb{R}} (1+ix)^{-\ell+k}(1-ix)^{-\ell-k} e(-xy)\,\mathrm{d}x.$$

The integral over $\mathbb{R}_+ \ni a$ is $\Gamma(|k|+\ell)$, while from [Gradshteyn and Ryzhik 2015, 3.384.9],

$$\int_{\mathbb{R}} (1+ix)^{-\ell+k}(1-ix)^{-\ell-k} e(-xy)\,\mathrm{d}x = \begin{cases} \frac{\pi^\ell |y|^{\ell-1}}{\Gamma(|k|+\ell)} W_{|k|,\ell-1/2}(4\pi|y|) & \text{if } \operatorname{sgn}(y) = \operatorname{sgn}(k), \\ 0 & \text{if } \operatorname{sgn}(y) = -\operatorname{sgn}(k). \end{cases}$$

This yields (54).

7B. ***The induced model.*** When ϕ is a shifted Hecke–Maaß cusp form $\varphi_{j,k}$, so that the associated irreducible representation π_∞ of $\mathrm{GL}_2(\mathbb{R})$ is a principal series representation, there is another natural model for π_∞ other than the Whittaker model $\mathcal{W}(\pi_\infty)$. This is the *induced model* of π_∞, which consists of smooth functions $f : \mathrm{GL}_2(\mathbb{R}) \to \mathbb{C}$ that satisfy

$$f\left(\begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix}\begin{pmatrix} y & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} z & 0 \\ 0 & z \end{pmatrix} g\right) = \operatorname{sgn}(y)^{\kappa_j}|y|^{1/2+ir_j} f(g)$$

for all $x \in \mathbb{R}$, $y, z \in \mathbb{R}^\times$, and $g \in \mathrm{GL}_2(\mathbb{R})$. Equivalently, the induced model consists of functions of the form

$$f(g) = \operatorname{sgn}(\det g)^{\kappa_j}|\det g|^{1/2+ir_j} \int_{\mathbb{R}^\times} \Phi((0,a)g)|a|^{1+2ir_j}\,\mathrm{d}^\times a \tag{55}$$

with $\Phi : \mathbb{R}^2 \to \mathbb{C}$ a Schwartz function. There is a bijection between the induced model and the Whittaker model via the map $f \mapsto W$ given by

$$W(g) = \lim_{N\to\infty} \int_{-N}^{N} f\left(\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} g\right) e(-x)\,\mathrm{d}x$$

[Jacquet and Langlands 1970, Lemma 2.5.13.1].

Taking Φ as in (51), we see that the element of the induced model f_ϕ associated to ϕ is such that

$$f_\phi\left(\begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix}\begin{pmatrix} y & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} z & 0 \\ 0 & z \end{pmatrix}\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}\right) = \operatorname{sgn}(y)^{\kappa_j}|y|^{1/2+ir_j} e^{2ki\theta} f_\phi \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{56}$$

for all $x \in \mathbb{R}$, $y, z \in \mathbb{R}^\times$, and $\theta \in \mathbb{R}$. From (55) together with the change of variables $a \mapsto \pi^{-1/2}|a|^{1/2}$, we have that

$$f_\phi \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \pi^{-1/2-ir_j}\Gamma\left(\tfrac{1}{2}+ir_j\right)\rho_j(1). \tag{57}$$

7C. *The Watson–Ichino triple product formula.* We now state an explicit form of the Watson–Ichino triple product formula, which relates the integral (47) to a triple product L-function and the square of the absolute value of certain local Whittaker functions and elements of the induced model.

Lemma 7.1 (Watson–Ichino triple product formula). *Let ϕ_i be Hecke cusp forms of weight $2k_i$ for which $k_1+k_2+k_3=0$ and such that ϕ_3 is a shifted Hecke–Maaß cusp form. Let W_1 and W_2 denote the local Whittaker functions associated to ϕ_1 and ϕ_2 respectively*, *as in* (52) *and* (54), *and let f_3 denote the element of the induced model associated to ϕ_3*, *as in* (56). *We have that*

$$\begin{aligned}\left|\int_{\mathrm{Z}(\mathbb{A}_\mathbb{Q})\,\mathrm{GL}_2(\mathbb{Q})\backslash\mathrm{GL}_2(\mathbb{A}_\mathbb{Q})} \widetilde{\phi}_1(g)\widetilde{\phi}_2(g)\widetilde{\phi}_3(g)\,\mathrm{d}g\right|^2 \\ = \frac{36}{\pi^2}L\left(\tfrac{1}{2},\phi_1\otimes\phi_2\otimes\phi_3\right)\left|\int_{\mathbb{R}^\times} W_1\begin{pmatrix} y & 0 \\ 0 & 1\end{pmatrix} W_2\begin{pmatrix} y & 0 \\ 0 & 1\end{pmatrix} f_3\begin{pmatrix} y & 0 \\ 0 & 1\end{pmatrix} |y|^{-1}\,\mathrm{d}^\times y\right|^2.\end{aligned} \tag{58}$$

Proof. This follows by combining the Watson–Ichino triple product formula in the form given in [Ichino 2008, Theorem 1.1] (compare [Watson 2002, Theorem 3]) together with the identities [Sarnak and Zhao 2019, Lemma 5] (compare [Michel and Venkatesh 2010, Lemma 3.4.2]) and [Waldspurger 1985, Proposition 6]. □

Remark 7.2. In place of the square of the absolute value of the integral on the right-hand side of (58), the Watson–Ichino triple product formula given in [Ichino 2008, Theorem 1.1] instead involves an integral of matrix coefficients. The utility of the induced model is that this integral of matrix coefficients may be expressed in terms of the simpler expression given in (58) [Sarnak and Zhao 2019, Lemma 5]. In turn, we shall shortly show that for our applications, this simpler expression can be explicitly evaluated in exactly the same way as in the Eisenstein setting in Lemmata 5.1 and 6.1. Thus utilizing the induced model allows us to express this integral in a way that is an exact analogue of the Eisenstein integrals in (44) and (45).

8. Maaß cusp form computation

We use the Watson–Ichino triple product formula to complete the next step of our main theorem, namely proving the desired bound for Hecke–Maaß cusp forms. The Watson–Ichino triple product formula allows us to relate an integral of a Hecke–Maaß cusp form and two shifted Hecke–Maaß cusp forms to the product of a ratio of L-functions and an integral involving Whittaker functions.

Lemma 8.1. *For any shifted Hecke–Maaß cusp form $\varphi_{\ell,k}$ of weight $2k \geq 0$ arising from a Hecke–Maaß cusp form φ_ℓ of weight 0 and spectral parameter r_ℓ, we have that*

$$\begin{aligned}\left|\int_{\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}} \varphi_j(z)\overline{\varphi_{j,k}(z)}\varphi_{\ell,k}(z)\,\mathrm{d}\mu(z)\right|^2 \\ = \frac{\pi^3}{2}\frac{L\left(\frac{1}{2},\varphi_\ell\right)L\left(\frac{1}{2},\mathrm{ad}\,\varphi_j\otimes\varphi_\ell\right)}{L(1,\mathrm{ad}\,\varphi_\ell)L(1,\mathrm{ad}\,\varphi_j)^2} \\ \times\left|\int_0^\infty \frac{W_{0,ir_j}(u)}{\Gamma\left(\frac{1}{2}+ir_j\right)}\left(\frac{W_{k,-ir_j}(u)}{\Gamma\left(\frac{1}{2}+k-ir_j\right)}+\frac{W_{-k,-ir_j}(u)}{\Gamma\left(\frac{1}{2}-k-ir_j\right)}\right)u^{-1/2+ir_\ell}\,\frac{\mathrm{d}u}{u}\right|^2. \end{aligned} \tag{59}$$

Proof. We apply the Watson–Ichino triple product formula (58), in conjunction with the classical-to-adèlic correspondence (47), in the case where the integrand is $\varphi_j\overline{\varphi_{j,k}}\varphi_{\ell,k}$. Thus we set

$$\phi_1 = \varphi_j, \quad \phi_2 = \overline{\varphi_{j,k}} \quad \text{and} \quad \phi_3 = \varphi_{\ell,k},$$

and we analyze the right-hand side of (58). We may factor the triple product L-function in (58) as

$$L\left(\tfrac{1}{2},\varphi_\ell\right)L\left(\tfrac{1}{2},\mathrm{ad}\,\varphi_j\otimes\varphi_\ell\right)$$

via [Gelbart and Jacquet 1978, (9.3) Theorem]. Note that both central L-values vanish unless φ_ℓ is even, which we assume without loss of generality is the case. We consider the remaining integral in (58). Recall that W_1 and W_2 are the local Whittaker functions associated to φ_j and $\overline{\varphi_{j,k}}$, while f_3 is the element of the induced model corresponding to the local Whittaker function W_3 for $\varphi_{\ell,k}$. From (52), we have that

$$W_1\begin{pmatrix} y & 0 \\ 0 & 1\end{pmatrix} = \mathrm{sgn}(y)^{\kappa_j}\rho_j(1)W_{0,ir_j}(4\pi|y|) \tag{60}$$

while comparing (21) and (49), we have that

$$W_2\begin{pmatrix} y & 0 \\ 0 & 1\end{pmatrix} = D_{k,-r_j}^{\mathrm{sgn}(y)}\,\mathrm{sgn}(y)^{\kappa_j}\rho_j(1)W_{\mathrm{sgn}(y)k,-ir_j}(4\pi|y|),$$

Finally, we have from (56) and (57) that

$$f_3\begin{pmatrix} y & 0 \\ 0 & 1\end{pmatrix} = \pi^{-1/2-ir_\ell}\Gamma\left(\tfrac{1}{2}+ir_\ell\right)\rho_\ell(1)|y|^{1/2+ir_\ell},$$

where the assumption that φ_ℓ is even means that we may omit $\operatorname{sgn}(y)^{\kappa_\ell}$. Inserting these formulæ and making the substitution $u = 4\pi|y|$, we deduce that

$$\begin{aligned}&\int_{\mathbb{R}^\times} W_1\begin{pmatrix} y & 0\\ 0 & 1\end{pmatrix} W_2\begin{pmatrix} y & 0\\ 0 & 1\end{pmatrix} f_3\begin{pmatrix} y & 0\\ 0 & 1\end{pmatrix} |y|^{-1}\,\mathrm{d}^\times y\\ &\qquad = 2(2\pi)^{-2ir_\ell}\Gamma\big(\tfrac{1}{2}+ir_\ell\big)\rho_\ell(1)\rho_j(1)^2\\ &\qquad\qquad \times \int_0^\infty W_{0,ir_j}(u)(D^+_{k,-r_j}W_{k,-ir_j}(u)+D^-_{k,-r_j}W_{-k,-ir_j}(u))u^{-1/2+ir_\ell}\,\frac{\mathrm{d}u}{u}.\end{aligned}$$

The desired identity now follows from the identities (22) for $D^\pm_{k,-r_j}$ (and noting that $\overline{\Gamma(z)} = \Gamma(\bar{z})$) and (20) for $\rho_\ell(1)$ and $\rho_j(1)^2$. □

Theorem 8.2. *For any $\delta > 0$ and $A > 0$, given a subconvex bound of the form*

$$L\big(\tfrac{1}{2}, \operatorname{ad}\varphi_1\otimes\varphi_2\big) \ll r_1^{1-4\delta} r_2^{2A}, \tag{61}$$

where φ_1, φ_2 are arbitrary Hecke–Maaß cusp forms with spectral parameters r_1, r_2, we have that

$$\begin{aligned}&\int_{\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}} \varphi_j(z)\overline{\varphi_{j,k}(z)}\varphi_{\ell,k}(z)\,\mathrm{d}\mu(z)\\ &\qquad\qquad \ll_{k,\varepsilon} r_j^{1/2-2\delta}\log r_j\, r_\ell^{A-1/4+\varepsilon}(2r_j+r_\ell)^{-1/4}(1+|2r_j-r_\ell|)^{-1/4}\end{aligned}$$

for any shifted Hecke–Maaß cusp form $\varphi_{\ell,k}$ of weight $2k$ and spectral parameter r_ℓ.

Proof. We consider only the positive weight case; the analogous bounds for the negative weight case follow by conjugational symmetry. We bound the expression (59). Via the assumption of the subconvex bound (61), the bound (32), and the convexity bound (36), the ratio of L-functions in (59) is $O_\varepsilon(r_j^{1-4\delta}(\log r_j)^2 r_\ell^{2A+1/2+\varepsilon})$. It remains to deal with the integral of Whittaker functions. In Corollary A.4, we show that this integral is $O_k(r_\ell^{-1/2}(2r_j+r_\ell)^{-1/4}(1+|2r_j-r_\ell|)^{-1/4})$. This yields the desired estimate. □

9. Holomorphic cusp form computation

We once more use the Watson–Ichino triple product formula in order to complete the final step of our main theorem, namely proving the desired bound for holomorphic or antiholomorphic Hecke cusp forms. The Watson–Ichino triple product formula allows us to relate an integral of a Hecke–Maaß cusp form, a shifted Hecke–Maaß cusp form, and a shifted holomorphic or antiholomorphic Hecke cusp form to the product of a ratio of L-functions and an integral involving Whittaker functions.

Lemma 9.1. *For any shifted holomorphic Hecke cusp form f_k of weight $2k > 0$ arising from a holomorphic Hecke cusp form F of weight $2\ell > 0$, we have that*

$$\left|\int_{\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}} \varphi_j(z)\overline{\varphi_{j,k}(z)}\, f_k(z)\,\mathrm{d}\mu(z)\right|^2 \\ = \frac{\pi^3}{2}\frac{L\big(\frac{1}{2},F\big)L\big(\frac{1}{2},\mathrm{ad}\,\varphi_j\otimes F\big)}{L(1,\mathrm{ad}\,F)L(1,\mathrm{ad}\,\varphi_j)^2} \\ \times\left|\int_0^\infty \frac{W_{0,ir_j}(u)}{\Gamma\big(\frac{1}{2}+ir_j\big)}\frac{W_{k,\ell-1/2}(u)}{\sqrt{\Gamma(k+\ell)\Gamma(k-\ell+1)}}u^{-1/2-ir_j}\,\frac{\mathrm{d}u}{u}\right|^2. \tag{62}$$

Proof. We apply the Watson–Ichino triple product formula (58), in conjunction with the classical-to-adèlic correspondence (47), in the case where the integrand is $\varphi_j\overline{\varphi_{j,k}}f_k$. Thus we set

$$\phi_1 = \varphi_j, \quad \phi_2 = f_k \quad \text{and} \quad \phi_3 = \overline{\varphi_{j,k}}.$$

We may again factor the triple product L-function in (58) as

$$L\big(\tfrac{1}{2},F\big)L\big(\tfrac{1}{2},\mathrm{ad}\,\varphi_j\otimes F\big)$$

via [Gelbart and Jacquet 1978, (9.3) Theorem]. Both central L-values vanish unless ℓ is even, which we assume without loss of generality is the case. We consider the remaining integral in (58). Here W_1 is once more as in (60). Next, W_2 is the local Whittaker function associated to f_k, so that from (54),

$$W_2\begin{pmatrix} y & 0\\ 0 & 1\end{pmatrix} = \begin{cases} C_{k,\ell}\rho_F(1)W_{k,\ell-\frac{1}{2}}(4\pi|y|) & \text{if } y>0,\\ 0 & \text{if } y<0,\end{cases}$$

noting that k is assumed to be positive. From (56) and (57), the element of the induced model associated to $\overline{\varphi_{j,k}}$ satisfies

$$f_3\begin{pmatrix} y & 0\\ 0 & 1\end{pmatrix} = \pi^{-1/2+ir_j}\Gamma\big(\tfrac{1}{2}-ir_j\big)\,\mathrm{sgn}(y)^{\kappa_j}\rho_j(1)|y|^{1/2-ir_j}.$$

Inserting these formulæ and making the substitution $u = 4\pi|y|$, we deduce that

$$\int_{\mathbb{R}^\times} W_1\begin{pmatrix} y & 0\\ 0 & 1\end{pmatrix} W_2\begin{pmatrix} y & 0\\ 0 & 1\end{pmatrix} f_3\begin{pmatrix} y & 0\\ 0 & 1\end{pmatrix}|y|^{-1}\,\mathrm{d}^\times y \\ = 2(2\pi)^{2ir_j}\Gamma\big(\tfrac{1}{2}-ir_j\big)C_{k,\ell}\rho_F(1)\rho_j(1)^2\int_0^\infty W_{0,ir_j}(u)W_{k,\ell-1/2}(u)u^{-1/2-ir_j}\,\frac{\mathrm{d}u}{u}.$$

The desired identity now follows from the identities (25) for $C_{k,\ell}$, (23) for $\rho_F(1)$, and (20) for $\rho_j(1)^2$. □

Theorem 9.2. *For any $\delta > 0$, given a subconvex bound of the form*

$$L\big(\tfrac{1}{2}, \operatorname{ad}\phi \otimes F\big) \ll_\ell r^{1-4\delta}, \tag{63}$$

where ϕ is an arbitrary Hecke–Maaß cusp form with spectral parameters r and F is a holomorphic Hecke cusp form of weight $2\ell > 0$, we have that

$$\int_{\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}} \varphi_j(z)\overline{\varphi_{j,k}(z)}\, f_k(z)\, \mathrm{d}\mu(z) \ll_{k,\ell} r_j^{-2\delta} \log r_j$$

for any shifted holomorphic or antiholomorphic Hecke cusp form f_k of weight $2k$ arising from a holomorphic Hecke cusp form F of weight $2\ell > 0$ for which $\ell \le |k|$.

Proof. We consider only the positive weight case; the analogous bounds for the negative weight case follow by conjugational symmetry. We bound the expression (62). Via the assumption of the subconvex bound (63) and the bound (32), the ratio of L-functions in (62) is $O_{\ell,\varepsilon}(r_j^{1-4\delta}(\log r_j)^2)$. It remains to deal with the integral of Whittaker functions. In Corollary A.7, we show that this integral is $O_{k,\ell}(r_j^{-1/2})$. This yields the desired estimate. □

10. Putting everything together

In this section, we prove Theorem 1.1.

Proof of Theorem 1.1. Let $a \in C_{c,K}^\infty(\mathrm{SL}_2(\mathbb{Z})\backslash \mathrm{SL}_2(\mathbb{R}))$. We recall from (18) that

$$\begin{aligned}
&\int_{\mathrm{SL}_2(\mathbb{Z})\backslash \mathrm{SL}_2(\mathbb{R})} a(z,\theta)\, \mathrm{d}\omega_j(z,\theta) \\
&\quad = \int_{\mathrm{SL}_2(\mathbb{Z})\backslash \mathrm{SL}_2(\mathbb{R})} a(z,\theta)\, \mathrm{d}\omega(z,\theta) \\
&\qquad + \frac{3}{\pi} \sum_{\ell=1}^{\infty} \sum_{k=-\infty}^{\infty} \langle a, \Phi_{\ell,k}\rangle \int_{\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}} \varphi_j(z)\overline{\varphi_{j,k}(z)}\varphi_{\ell,k}(z)\, \mathrm{d}\mu(z) \\
&\qquad + \frac{3}{\pi} \sum_{\ell=1}^{\infty} \sum_{F\in\mathcal{H}_\ell} \sum_{\substack{k=-\infty \\ |k|\ge\ell}}^{\infty} \langle a, \Psi_{F,k}\rangle \int_{\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}} \varphi_j(z)\overline{\varphi_{j,k}(z)}\, f_k(z)\, \mathrm{d}\mu(z) \\
&\qquad + \frac{3}{4\pi^2} \sum_{k=-\infty}^{\infty} \int_{-\infty}^{\infty} \big\langle a, \widetilde{E}_{2k}\big(\cdot,\cdot,\tfrac{1}{2}+it\big)\big\rangle \\
&\qquad\qquad \times \int_{\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}} \varphi_j(z)\overline{\varphi_{j,k}(z)} E_{2k}\big(z,\tfrac{1}{2}+it\big)\, \mathrm{d}\mu(z)\, \mathrm{d}t.
\end{aligned}$$

Since $\Phi_{\ell,k}$, $\Psi_{F,k}$, and $\widetilde{E}_{2k}\big(\cdot,\cdot,\frac{1}{2}+it\big)$ are eigenfunctions of the Casimir operator Ω, we have that for any nonnegative integer A,

$$\begin{aligned}\Omega^A \Phi_{\ell,k} &= \big(\tfrac{1}{4}+r_\ell^2\big)^A \Phi_{\ell,k},\\ \Omega^A \Psi_{F,k} &= (\ell(1-\ell))^A \Psi_{F,k},\\ \Omega^A \widetilde{E}_{2k}\big(\cdot,\cdot,\tfrac{1}{2}+it\big) &= \big(\tfrac{1}{4}+t^2\big)^A \widetilde{E}_{2k}\big(\cdot,\cdot,\tfrac{1}{2}+it\big).\end{aligned}$$

By the linearity of the inner product together with the fact that the Casimir operator is self-adjoint with respect to the inner product, we deduce that

$$\begin{aligned}\langle a, \Phi_{\ell,k}\rangle &= \big(\tfrac{1}{4}+r_\ell^2\big)^{-A} \langle \Omega^A a, \Phi_{\ell,k}\rangle,\\ \langle a, \Psi_{F,k}\rangle &= (\ell(1-\ell))^{-A} \langle \Omega^A a, \Psi_{F,k}\rangle,\\ \big\langle a, \widetilde{E}_{2k}\big(\cdot,\cdot,\tfrac{1}{2}+it\big)\big\rangle &= \big(\tfrac{1}{4}+t^2\big)^{-A} \big\langle \Omega^A a, \widetilde{E}_{2k}\big(\cdot,\cdot,\tfrac{1}{2}+it\big)\big\rangle.\end{aligned}$$

Moreover, since a is K-finite, there exists a nonnegative integer M for which

$$\langle a, \Phi_{\ell,k}\rangle = \langle a, \Psi_{F,k}\rangle = \big\langle a, \widetilde{E}_{2k}\big(\cdot,\cdot,\tfrac{1}{2}+it\big)\big\rangle = 0$$

whenever $|k| > M$. From Theorems 6.2, 8.2, and 9.2, we deduce that

$$\begin{aligned}&\int_{\mathrm{SL}_2(\mathbb{Z})\backslash \mathrm{SL}_2(\mathbb{R})} a(z,\theta)\,\mathrm{d}\omega_j(z,\theta) - \int_{\mathrm{SL}_2(\mathbb{Z})\backslash \mathrm{SL}_2(\mathbb{R})} a(z,\theta)\,\mathrm{d}\omega(z,\theta)\\ &\quad\ll_{M,\varepsilon} r_j^{1/2-2\delta}\log r_j \sum_{\ell=1}^{\infty}\sum_{k=-M}^{M} \big|\big\langle \Omega^{\lceil (A+1)/2\rceil} a, \Phi_{\ell,k}\big\rangle\big| r_\ell^{-5/4+\varepsilon}\\ &\qquad\qquad\qquad\qquad\qquad\qquad \times (2r_j+r_\ell)^{-1/4}(1+|2r_j-r_\ell|)^{-1/4}\\ &\qquad + r_j^{-2\delta}\log r_j \sum_{\ell=1}^{M}\sum_{F\in\mathcal{H}_\ell}\sum_{\substack{k=-M\\ |k|\geq \ell}}^{M} |\langle \Omega^{\lceil (A+1)/2\rceil} a, \Psi_{F,k}\rangle| \ell^{-A-1}\\ &\qquad + r_j^{1/2-2\delta}\log r_j \sum_{k=-M}^{M}\int_{-\infty}^{\infty} \big|\big\langle \Omega^{\lceil (A+1)/2\rceil} a, \widetilde{E}_{2k}\big(\cdot,\cdot,\tfrac{1}{2}+it\big)\big\rangle\big|\\ &\qquad\qquad\qquad \times (1+|t|)^{-5/4+\varepsilon}(2r_j+|t|)^{-1/4}(1+|2r_j-|t||)^{-1/4}\,\mathrm{d}t.\end{aligned}$$

Weyl's law [Risager 2004, Theorem 2] implies that $\#\{\ell\in\mathbb{N} : T-1 < r_\ell \leq T\} \ll T$ for $T \geq 1$. Thus by subdividing the sum over $\ell \in \mathbb{N}$ into sums for which $r_\ell \in (T-1, T]$ for each $T \in \mathbb{N}$, we find that

$$\sum_{\ell=1}^{\infty} r_\ell^{-5/2+\varepsilon}(2r_j+r_\ell)^{-1/2}(1+|2r_j-r_\ell|)^{-1/2} \ll \frac{1}{r_j}.$$

Similarly,

$$\int_{-\infty}^{\infty} (1+|t|)^{-5/2+\varepsilon}(2r_j+|t|)^{-1/2}(1+|2r_j-|t||)^{-1/2}\,\mathrm{d}t \ll \frac{1}{r_j}.$$

Thus by the Cauchy–Schwarz inequality and Bessel's inequality (bearing in mind Parseval's identity (14)), we deduce that

$$\int_{\mathrm{SL}_2(\mathbb{Z})\backslash \mathrm{SL}_2(\mathbb{R})} a(z,\theta)\,\mathrm{d}\omega_j(z,\theta) - \int_{\mathrm{SL}_2(\mathbb{Z})\backslash \mathrm{SL}_2(\mathbb{R})} a(z,\theta)\,\mathrm{d}\omega(z,\theta) \\ \ll_M \|\Omega^{\lceil (A+1)/2\rceil} a\|_{L^2(\mathrm{SL}_2(\mathbb{Z})\backslash \mathrm{SL}_2(\mathbb{R}))} r_j^{-2\delta} \log r_j, \quad (64)$$

which completes the proof. □

Remark 10.1. Theorem 1.1 is proven for functions $a : \mathrm{SL}_2(\mathbb{Z})\backslash \mathrm{SL}_2(\mathbb{R}) \to \mathbb{C}$ that are finite linear combinations of even weight smooth compactly supported functions. In order to remove the condition that a be a finite linear combination of even weight functions, we would require bounds for the integral (15) that are uniform not only in r_j and r_ℓ but additionally uniform in k; we would also similarly require such uniform bounds for the integrals (16) and (17). To prove such uniform bounds would require stronger bounds for certain hypergeometric functions than the weaker bounds we derive in Lemma A.3 and Corollary A.7 below, as we discuss in Remark A.8.

Appendix: Whittaker integral computations

AA. ***Special functions.*** We compute integrals of Whittaker functions by expressing them in terms of generalized hypergeometric functions, as defined in [Slater 1966, Chapter 2]. A generalized hypergeometric function is defined, wherever it converges, as a series

$${}_pF_q\left(\begin{matrix} a_1,\dots,a_p \\ b_1,\dots,b_q\end{matrix}; z\right) := \sum_{m=0}^{\infty} \frac{(a_1)_m \cdots (a_p)_m}{(b_1)_m \cdots (b_q)_m} \frac{z^m}{m!}. \quad (1)$$

Here $(b)_m := b(b+1)\cdots(b+m-1)$ and $(b)_0 := 1$ for all $b \in \mathbb{C}$, so that

$$(b)_m = \begin{cases} \frac{\Gamma(b+m)}{\Gamma(b)} & \text{if } b \text{ is not a nonpositive integer,} \\ (-1)^m \frac{\Gamma(1-b)}{\Gamma(1-b-m)} & \text{if } b \text{ is a nonpositive integer and } m \le -b, \\ 0 & \text{if } b \text{ is a nonpositive integer and } m > -b. \end{cases} \quad (2)$$

To bound hypergeometric functions, we must therefore bound gamma functions. We do this via Stirling's formula, which states that for $s \in \mathbb{C}$ with $\Re(s) > \delta$ with $\delta > 0$,

$$\Gamma(s) = \sqrt{2\pi}\, s^{s-1/2} e^{-s}\left(1 + O_\delta\left(\frac{1}{|s|}\right)\right).$$

We use this in the following form: for $s = \sigma + i\tau$ with $\sigma > 0$,

$$|\Gamma(\sigma + i\tau)| \asymp_\sigma (1+|\tau|)^{\sigma-1/2} e^{-(\pi/2)|\tau|}. \quad (3)$$

AB. ***Nonholomorphic case.*** We seek to provide an upper bound for an integral of the form

$$I_k(\alpha,\beta,\gamma)=\int_0^\infty \frac{W_{0,i\alpha}(y)}{\Gamma\bigl(\frac{1}{2}+i\alpha\bigr)}\biggl(\frac{W_{k,i\beta}(y)}{\Gamma\bigl(\frac{1}{2}+k+i\beta\bigr)}+\frac{W_{-k,i\beta}(y)}{\Gamma\bigl(\frac{1}{2}-k+i\beta\bigr)}\biggr)y^{-1/2+i\gamma}\,\frac{\mathrm{d}y}{y},$$

where $k\in\mathbb{Z}$ and $\alpha,\beta,\gamma\in\mathbb{R}$. This can be expressed in terms of gamma functions and a terminating hypergeometric function.

Lemma A.1 [Jakobson 1997, (27)]. *For $k\in\mathbb{Z}$ and $\alpha,\beta,\gamma\in\mathbb{R}$, we have that*

$$\begin{aligned}I_k(\alpha,\beta,\gamma)&=\frac{(-1)^k4^{i\gamma}}{2\pi}\frac{\prod_{\epsilon_1,\epsilon_2\in\{\pm1\}}\Gamma\bigl(\frac{1}{4}+\frac{i}{2}(\epsilon_1\alpha+\epsilon_2\beta+\gamma)\bigr)}{\Gamma\bigl(\frac{1}{2}+i\alpha\bigr)\Gamma\bigl(\frac{1}{2}+i\beta\bigr)\Gamma\bigl(\frac{1}{2}+i\gamma\bigr)}\\&\qquad\times{}_4F_3\biggl(\begin{matrix}-k,\,k,\,\frac{1}{4}+\frac{i}{2}(-\alpha+\beta+\gamma),\,\frac{1}{4}+\frac{i}{2}(\alpha+\beta+\gamma)\\ \frac{1}{2},\,\frac{1}{2}+i\beta,\,\frac{1}{2}+i\gamma\end{matrix};1\biggr).\end{aligned}\tag{4}$$

To obtain uniform bounds for the expression (4), we first deal with the ratio of gamma functions.

Lemma A.2. *For $r,t\in\mathbb{R}$, we have that*

$$\begin{aligned}&\frac{\Gamma\bigl(\frac{1}{4}+\frac{i(2r+t)}{2}\bigr)\Gamma\bigl(\frac{1}{4}+\frac{it}{2}\bigr)^2\Gamma\bigl(\frac{1}{4}+\frac{i(-2r+t)}{2}\bigr)}{\Gamma\bigl(\frac{1}{2}+ir\bigr)\Gamma\bigl(\frac{1}{2}-ir\bigr)\Gamma\bigl(\frac{1}{2}+it\bigr)}\\&\ll\begin{cases}(1+|t|)^{-1/2}(1+|2r+t|)^{-1/4}(1+|2r-t|)^{-1/4} & \textit{if } |t|\le 2|r|,\\(1+|t|)^{-1/2}(1+|2r+t|)^{-1/4}(1+|2r-t|)^{-1/4}e^{-(\pi/2)(|t|-2|r|)} & \textit{if } |t|\ge 2|r|.\end{cases}\end{aligned}$$

Proof. This follows from Stirling's formula (3). □

Next, we bound the hypergeometric function in (4).

Lemma A.3. *For $k\in\mathbb{Z}$ and $r,t\in\mathbb{R}$, we have that*

$${}_4F_3\biggl(\begin{matrix}-k,\,k,\,\frac{1}{4}+\frac{i(-2r+t)}{2},\,\frac{1}{4}+\frac{it}{2}\\ \frac{1}{2},\,\frac{1}{2}-ir,\,\frac{1}{2}+it\end{matrix};1\biggr)\ll_k 1+\biggl(\frac{1+|2r-t|}{1+|r|}\biggr)^{|k|}.$$

Proof. By (1) and (2), the left-hand side is

$$\begin{aligned}&\sum_{m=0}^{|k|}\frac{\sqrt{\pi}\,|k|(-1)^m(|k|+m-1)!}{(|k|-m)!\,\Gamma\bigl(\frac{1}{2}+m\bigr)m!}\\&\qquad\times\frac{\Gamma\bigl(m+\frac{1}{4}+\frac{i(-2r+t)}{2}\bigr)\Gamma\bigl(m+\frac{1}{4}+\frac{it}{2}\bigr)\Gamma\bigl(\frac{1}{2}-ir\bigr)\Gamma\bigl(\frac{1}{2}+it\bigr)}{\Gamma\bigl(\frac{1}{4}+\frac{i(-2r+t)}{2}\bigr)\Gamma\bigl(\frac{1}{4}+\frac{it}{2}\bigr)\Gamma\bigl(\frac{1}{2}+m-ir\bigr)\Gamma\bigl(\frac{1}{2}+m+it\bigr)}.\end{aligned}$$

By Stirling's formula (3), each summand is

$$\ll_k\biggl(\frac{1+|2r-t|}{1+|r|}\biggr)^m.$$

This yields the desired bounds. □

Combining Lemmata A.2 and A.3, we deduce the following bounds; these bounds are not sharp but are more than sufficient for our purposes.

Corollary A.4. *For $k \in \mathbb{Z}$ and $r, t \in \mathbb{R}$,*

$$I_k(r,-r,t) \ll_k (1+|t|)^{-1/2}(1+|2r+t|)^{-1/4}(1+|2r-t|)^{-1/4}.$$

AC. *Holomorphic case.* Here we instead seek to provide an upper bound for an integral of the form

$$I_{k,\ell}(r) := \int_0^\infty \frac{W_{0,ir}(u)}{\Gamma\bigl(\frac{1}{2}+ir\bigr)} \frac{W_{k,\ell-1/2}(u)}{\sqrt{\Gamma(k+\ell)\Gamma(k-\ell+1)}} u^{-1/2-ir}\,\frac{\mathrm{d}u}{u},$$

where $k, \ell \in \mathbb{N}$ are positive integers for which $k \geq \ell$ and $r \in \mathbb{R}$.

Lemma A.5. *For $k, \ell \in \mathbb{N}$ for which $k \geq \ell$ and for $r \in \mathbb{R}$, we have that*

$$\begin{aligned} I_{k,\ell}(r) = (-1)^{k-\ell}\sqrt{\frac{\pi}{2}} \frac{\sqrt{\Gamma(k+\ell)\Gamma(k-\ell+1)}}{\Gamma\bigl(\frac{1}{2}+ir\bigr)} \\ \times \sum_{m=0}^{k-\ell} \frac{(-1)^m(\ell+m-1)!\,\Gamma(\ell+m-2ir)}{(k-\ell-m)!(2\ell+m-1)!\,\Gamma\bigl(\frac{1}{2}+\ell+m-ir\bigr)m!}. \end{aligned} \tag{5}$$

Proof. We use the fact that

$$\begin{aligned} &W_{k,\ell-1/2}(u) \\ &\quad = (-1)^{k-\ell}(k-\ell)!\,e^{-u/2}u^{\ell}L_{k-\ell}^{(2\ell-1)}(u) \\ &\quad = (-1)^{k-\ell}(k-\ell)!(k+\ell-1)!\sum_{m=0}^{k-\ell}(-1)^m/(k-\ell-m)!(2\ell+m-1)!m!u^{\ell+m}e^{-u/2} \end{aligned}$$

from [Gradshteyn and Ryzhik 2015, (8.970.1) and (9.237.3)], where $L_n^{(\alpha)}$ denotes the associated Laguerre polynomial, together with the identity

$$\int_0^\infty W_{0,ir}(u)e^{-u/2}u^{\ell+m-1/2-ir}\,\frac{\mathrm{d}u}{u} = \frac{(\ell+m-1)!\,\Gamma(\ell+m-2ir)}{\Gamma\bigl(\frac{1}{2}+\ell+m-ir\bigr)}$$

from [Gradshteyn and Ryzhik 2015, (7.621.11)], in order to obtain the desired identity. □

Remark A.6. Via (1) and (2), we may write $I_{k,\ell}(r)$ in the form

$$(-1)^{k-\ell}\sqrt{\frac{\pi}{2}} \frac{\sqrt{\Gamma(k+\ell)}\,\Gamma(\ell)\Gamma(\ell-2ir)}{\sqrt{\Gamma(k-\ell+1)}\,\Gamma(2\ell)\Gamma\bigl(\frac{1}{2}+ir\bigr)\Gamma\bigl(\frac{1}{2}+\ell-ir\bigr)} {}_3F_2\left(\begin{matrix}\ell-k,\ell,\ell-2ir \\ 2\ell,\frac{1}{2}+\ell-ir\end{matrix};1\right).$$

One can show that this can alternatively be written as

$$(-1)^{k-\ell}\sqrt{\frac{\pi}{2}}\frac{\Gamma(\ell)}{\sqrt{\Gamma(k+\ell)\Gamma(k-\ell+1)}}\frac{\Gamma\left(\frac{1}{2}+k+ir\right)\Gamma(\ell-2ir)}{\Gamma\left(\frac{1}{2}+ir\right)\Gamma\left(\frac{1}{2}+\ell+ir\right)\Gamma\left(\frac{1}{2}+\ell-ir\right)} \times {}_3F_2\left(\begin{matrix}\ell-k,\, \frac{1}{2}+ir,\, \frac{1}{2}-ir \\ \frac{1}{2}+\ell+ir,\, \frac{1}{2}+\ell-ir\end{matrix};1\right).$$

However, we do not make use of these identities.

We now bound $I_{k,\ell}(r)$.

Corollary A.7. *For $k, \ell \in \mathbb{N}$ for which $k \geq \ell$ and $r \in \mathbb{R}$, we have that*

$$I_{k,\ell}(r) \ll_{k,\ell} (1+|r|)^{-1/2}.$$

Proof. We simply bound each summand in (5) via Stirling's formula (3). □

Remark A.8. As highlighted in Remark 10.1, it would be desirable to prove bounds for $I_k(r, -r, t)$ that are uniform not only with respect to r and t but also with respect to k. Similarly, it would be desirable to prove bounds for $I_{k,\ell}(r)$ that are uniform not only with respect to r but also with respect to k and ℓ. A closer examination of the method of proofs of Corollaries A.4 and A.7 shows that these methods can be used to give bounds that grow exponentially with k, which is insufficient for our needs. Were we to fix every variable except k, then the methods in [Fields 1965] can be used to obtain polynomial bounds for both of $I_k(r, -r, t)$ and $I_{k,\ell}(r)$ solely in the k-aspect; unfortunately, however, this is also insufficient for our needs.

Acknowledgements

The authors participated in, and conducted this research through, the 2023 UVA REU in Number Theory. They are grateful for the support of grants from Jane Street Capital, the National Science Foundation (DMS-2002265 and DMS-2147273), the National Security Agency (H98230-23-1-0016), and the Templeton World Charity Foundation. The second author was additionally supported by the National Science Foundation (grant DMS-2302079) and by the Simons Foundation (award 965056).

The authors would like to thank Maximiliano Sanchez Garza for his guidance, Asaf Katz and Jesse Thorner for useful feedback, and the anonymous referees for their thorough readings of this paper. Finally, they are grateful to Ken Ono for organizing the 2023 UVA REU.

References

[Anantharaman 2010] N. Anantharaman, "Eigenfunctions of the Laplacian on negatively curved manifolds: a semiclassical approach", pp. 389–438 in *Homogeneous flows, moduli spaces and* *arithmetic*, edited by M. L. Einsiedler et al., Clay Math. Proc. **10**, Amer. Math. Soc., Providence, RI, 2010. MR Zbl

[Bump 1997] D. Bump, *Automorphic forms and representations*, Cambridge Studies in Advanced Mathematics **55**, Cambridge University Press, 1997. MR Zbl

[De Bièvre 2001] S. De Bièvre, "Quantum chaos: a brief first visit", pp. 161–218 in *Second summer school in analysis and mathematical physics* (Cuernavaca, 2000), edited by S. Pérez-Esteva and C. Villegas-Blas, Contemp. Math. **289**, Amer. Math. Soc., Providence, RI, 2001. MR Zbl

[Duke et al. 2002] W. Duke, J. B. Friedlander, and H. Iwaniec, "The subconvexity problem for Artin L-functions", *Invent. Math.* **149**:3 (2002), 489–577. MR Zbl

[Dyatlov 2022] S. Dyatlov, "Around quantum ergodicity", *Ann. Math. Qué.* **46**:1 (2022), 11–26. MR Zbl

[Fields 1965] J. L. Fields, "Asymptotic expansions of a class of hypergeometric polynomials with respect to the order, III", *J. Math. Anal. Appl.* **12** (1965), 593–601. MR Zbl

[Gelbart and Jacquet 1978] S. Gelbart and H. Jacquet, "A relation between automorphic representations of GL(2) and GL(3)", *Ann. Sci. École Norm. Sup.* (4) **11**:4 (1978), 471–542. MR Zbl

[Goldfeld and Hundley 2011] D. Goldfeld and J. Hundley, *Automorphic representations and L-functions for the general linear group, I*, Cambridge Studies in Advanced Mathematics **129**, Cambridge University Press, 2011. MR Zbl

[Goldfeld et al. 1994] D. Goldfeld, J. Hoffstein, and D. Lieman, "An effective zero-free region", (1994). Appendix to J. Hoffstein and P. Lockhard P, "Coefficients of Maass forms and the Siegel zero", Ann. of Math. (2) **140**:1 (1994), 161–181. MR Zbl

[Good 1982] A. Good, "The square mean of Dirichlet series associated with cusp forms", *Mathematika* **29**:2 (1982), 278–295. MR Zbl

[Gradshteyn and Ryzhik 2015] I. S. Gradshteyn and I. M. Ryzhik, *Table of integrals*, *series*, *and products*, 8th ed., Elsevier, Amsterdam, 2015. MR Zbl

[Hassell 2010] A. Hassell, "Ergodic billiards that are not quantum unique ergodic", *Ann. of Math.* (2) **171**:1 (2010), 605–618. MR Zbl

[Hassell 2011] A. Hassell, "What is quantum unique ergodicity?", *Austral. Math. Soc. Gaz.* **38**:3 (2011), 158–167. MR Zbl

[Humphries and Nordentoft 2022] P. Humphries and A. C. Nordentoft, "Sparse equidistribution of geometric invariants of real quadratic fields", preprint, 2022. Zbl arXiv 2211.05890v1

[Ichino 2008] A. Ichino, "Trilinear forms and the central values of triple product L-functions", *Duke Math. J.* **145**:2 (2008), 281–307. MR Zbl

[Iwaniec 2002] H. Iwaniec, *Spectral methods of automorphic forms*, 2nd ed., Graduate Studies in Mathematics **53**, American Mathematical Society, Providence, RI, 2002. MR Zbl

[Iwaniec and Kowalski 2004] H. Iwaniec and E. Kowalski, *Analytic number theory*, American Mathematical Society Colloquium Publications **53**, American Mathematical Society, Providence, RI, 2004. MR Zbl

[Jacquet and Langlands 1970] H. Jacquet and R. P. Langlands, *Automorphic forms on* GL(2), Lecture Notes in Mathematics **Vol. 114**, Springer, 1970. MR Zbl

[Jakobson 1994] D. Jakobson, "Quantum unique ergodicity for Eisenstein series on $\mathrm{PSL}_2(\mathbb{Z})\backslash\mathrm{PSL}_2(\mathbb{R})$", *Ann. Inst. Fourier* (*Grenoble*) **44**:5 (1994), 1477–1504. MR Zbl

[Jakobson 1997] D. Jakobson, "Equidistribution of cusp forms on $\mathrm{PSL}_2(\mathbb{Z})\backslash\mathrm{PSL}_2(\mathbb{R})$", *Ann. Inst. Fourier* (*Grenoble*) **47**:3 (1997), 967–984. MR Zbl

[Katok 1992] S. Katok, *Fuchsian groups*, University of Chicago Press, 1992. MR Zbl

[Khan and Young 2023] R. Khan and M. P. Young, "Moments and hybrid subconvexity for symmetric-square L-functions", *J. Inst. Math. Jussieu* **22**:5 (2023), 2029–2073. MR Zbl

[Lang 1985] S. Lang, $\mathrm{SL}_2(\mathbf{R})$, Graduate Texts in Mathematics **105**, Springer, 1985. MR

[Lau and Wu 2006] Y.-K. Lau and J. Wu, "A density theorem on automorphic L-functions and some applications", *Trans. Amer. Math. Soc.* **358**:1 (2006), 441–472. MR Zbl

[Lazutkin 1993] V. F. Lazutkin, *KAM theory and semiclassical approximations to eigenfunctions*, Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)] **24**, Springer, 1993. MR Zbl

[Li 2010] X. Li, "Upper bounds on L-functions at the edge of the critical strip", *Int. Math. Res. Not.* **2010**:4 (2010), 727–755. MR Zbl

[Lindenstrauss 2001] E. Lindenstrauss, "On quantum unique ergodicity for $\Gamma\backslash\mathbb{H}\times\mathbb{H}$", *Internat. Math. Res. Notices* 17 (2001), 913–933. MR Zbl

[Lindenstrauss 2006] E. Lindenstrauss, "Invariant measures and arithmetic quantum unique ergodicity", *Ann. of Math.* (2) **163**:1 (2006), 165–219. MR Zbl

[Luo and Sarnak 1995] W. Z. Luo and P. Sarnak, "Quantum ergodicity of eigenfunctions on $\mathrm{PSL}_2(\mathbf{Z})\backslash\mathbf{H}^2$", *Inst. Hautes Études Sci. Publ. Math.* 81 (1995), 207–237. MR Zbl

[Marklof 2006] J. Marklof, "Arithmetic quantum chaos", pp. 212–221 in *Encyclopedia of mathematical physics*, vol. 1, edited by J.-P. Fran¸coise et al., Academic Press, 2006. MR Zbl

[Marshall 2014] S. Marshall, "Triple product L-functions and quantum chaos on $SL(2,\mathbb{C})$", *Israel J. Math.* **200**:1 (2014), 423–448. MR Zbl

[Meurman 1990] T. Meurman, "On the order of the Maass L-function on the critical line", pp. 325–354 in *Number theory*, *I* (Budapest, 1987), edited by K. Gy˝ ory and G. Halász, Colloq. Math. Soc. János Bolyai **51**, North-Holland, Amsterdam, 1990. MR Zbl

[Michel and Venkatesh 2010] P. Michel and A. Venkatesh, "The subconvexity problem for GL_2", *Publ. Math. Inst. Hautes Études Sci.* 111 (2010), 171–271. MR Zbl

[Nelson 2023] P. D. Nelson, "Bounds for standard L-functions", preprint, 2023. Zbl arXiv 2109.15230v2

[Nonnenmacher 2013] S. Nonnenmacher, "Anatomy of quantum chaotic eigenstates", pp. 193–238 in *Chaos*, edited by B. Duplantier et al., Prog. Math. Phys. **66**, Springer, 2013. MR Zbl

[Risager 2004] M. S. Risager, "Asymptotic densities of Maass newforms", *J. Number Theory* **109**:1 (2004), 96–119. MR Zbl

[Roelcke 1966] W. Roelcke, "Das Eigenwertproblem der automorphen Formen in der hyperbolischen Ebene, I", *Math. Ann.* **167**:4 (1966), 292–337. MR Zbl

[Rudnick and Sarnak 1994] Z. Rudnick and P. Sarnak, "The behaviour of eigenstates of arithmetic hyperbolic manifolds", *Comm. Math. Phys.* **161**:1 (1994), 195–213. MR Zbl

[Sarnak 2011] P. Sarnak, "Recent progress on the quantum unique ergodicity conjecture", *Bull. Amer. Math. Soc.* (*N.S.*) **48**:2 (2011), 211–228. MR Zbl

[Sarnak and Zhao 2019] P. Sarnak and P. Zhao, "The quantum variance of the modular surface", *Ann. Sci. Éc. Norm. Supér.* (4) **52**:5 (2019), 1155–1200. MR Zbl

[Shnirelman 1974] A. I. Shnirelman, "Ergodic properties of eigenfunctions", *Uspehi Mat. Nauk* **29**:6(180) (1974), 181–182. In Russian. MR Zbl

[Slater 1966] L. J. Slater, *Generalized hypergeometric functions*, Cambridge University Press, 1966. MR Zbl

[Soundararajan 2010] K. Soundararajan, "Quantum unique ergodicity for $\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}$", *Ann. of Math.* (2) **172**:2 (2010), 1529–1538. MR Zbl

[Colin de Verdière 1985] Y. Colin de Verdière, "Ergodicité et fonctions propres du laplacien", *Comm. Math. Phys.* **102**:3 (1985), 497–502. MR Zbl

[Waldspurger 1985] J.-L. Waldspurger, "Sur les valeurs de certaines fonctions L automorphes en leur centre de symétrie", *Compositio Math.* **54**:2 (1985), 173–242. MR Zbl

[Watson 2002] T. C. Watson, *Rankin triple products and quantum chaos*, Ph.D. thesis, Princeton University, 2002, available at https://www.proquest.com/docview/252081445. MR Zbl

[Whittaker and Watson 1996] E. T. Whittaker and G. N. Watson, *A course of modern analysis*, Cambridge University Press, 1996. MR Zbl

[Wu 2017] H. Wu, "A note on spectral analysis in automorphic representation theory for GL_2, I", *Int. J. Number Theory* **13**:10 (2017), 2717–2750. MR Zbl

[Young 2016] M. P. Young, "The quantum unique ergodicity conjecture for thin sets", *Adv. Math.* **286** (2016), 958–1016. MR Zbl

[Zelditch 1987] S. Zelditch, "Uniform distribution of eigenfunctions on compact hyperbolic surfaces", *Duke Math. J.* **55**:4 (1987), 919–941. MR Zbl

[Zelditch 1991] S. Zelditch, "Mean Lindelöf hypothesis and equidistribution of cusp forms and Eisenstein series", *J. Funct. Anal.* **97**:1 (1991), 1–49. MR Zbl

[Zelditch 2006] S. Zelditch, "Quantum ergodicity and mixing of eigenfunctions", pp. 183–196 in *Encyclopedia of mathematical physics*, vol. 4, edited by J.-P. Fran¸coise et al., Academic Press, 2006. MR Zbl

[Zelditch 2010] S. Zelditch, "Recent developments in mathematical quantum chaos", pp. 115–204 in *Current developments in mathematics*, 2009, edited by D. Jerison et al., International Press, Somerville, MA, 2010. MR Zbl

[Zelditch 2019] S. Zelditch, "Mathematics of quantum chaos in 2019", *Notices Amer. Math. Soc.* **66**:9 (2019), 1412–1422. MR Zbl

Received 23 Feb 2024. Revised 5 Sep 2024.

ANKIT BISAIN:

ankitb12@mit.edu
Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA, United States

PETER HUMPHRIES:

pclhumphries@gmail.com
Department of Mathematics, University of Virginia, Charlottesville, VA, United States

ANDREI MANDELSHTAM:

andman@stanford.edu
Department of Mathematics, Stanford University, Stanford, CA, United States

NOAH WALSH:

nwalsh21@mit.edu
Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA, United States

XUN WANG:

stevenwang2029@u.northwestern.edu
Department of Mathematics, Northwestern University, Evanston, IL, United States

Guidelines for Authors

Authors may submit manuscripts in PDF format on-line at the submission page.

Originality. Submission of a manuscript acknowledges that the manuscript is original and and is not, in whole or in part, published or under consideration for publication elsewhere. It is understood also that the manuscript will not be submitted elsewhere while under consideration for publication in this journal.

Language. Articles are usually in English or French, but articles written in other languages are welcome.

Required items. A brief abstract of about 150 words or less must be included. It should be self-contained and not refer to bibliography keys. If the article is not in English, two versions of the abstract must be included, one in the language of the article and one in English. Also required are keywords and a Mathematics Subject Classification for the article, and, for each author, affiliation (if appropriate) and email address.

Format. Authors are encouraged to use LaTeX and the standard amsart class, but submissions in other varieties of TeX, and exceptionally in other formats, are acceptable. Initial uploads should normally be in PDF format; after the refereeing process we will ask you to submit all source material.

References. Bibliographical references should be complete, including article titles and page ranges. All references in the bibliography should be cited in the text. The use of BibTeX is preferred but not required. Tags will be converted to the house format, however, for submission you may use the format of your choice. Links will be provided to all literature with known web locations and authors are encouraged to provide their own links in addition to those supplied in the editorial process.

Figures. Figures must be of publication quality. After acceptance, you will need to submit the original source files in vector graphics format for all diagrams in your manuscript: vector EPS or vector PDF files are the most useful.

Most drawing and graphing packages — Mathematica, Adobe Illustrator, Corel Draw, MATLAB, etc. — allow the user to save files in one of these formats. Make sure that what you are saving is vector graphics and not a bitmap. If you need help, please write to graphics@msp.org with as many details as you can about how your graphics were generated.

Bundle your figure files into a single archive (using zip, tar, rar or other format of your choice) and upload on the link you been provided at acceptance time. Each figure should be captioned and numbered so that it can float. Small figures occupying no more than three lines of vertical space can be kept in the text (“the curve looks like this:”). It is acceptable to submit a manuscript with all figures at the end, if their placement is specified in the text by means of comments such as “Place Figure 1 here”. The same considerations apply to tables.

White Space. Forced line breaks or page breaks should not be inserted in the document. There is no point in your trying to optimize line and page breaks in the original manuscript. The manuscript will be reformatted to use the journal’s preferred fonts and layout.

Proofs. Page proofs will be made available to authors (or to the designated corresponding author) at a Web site in PDF format. Failure to acknowledge the receipt of proofs or to return corrections within the requested deadline may cause publication to be postponed.

ESSENTIAL NUMBER THEORY

2024 vol. 3 no. 1