Modular elliptic curves over real abelian fields and the generalized Fermat equation $x^{2\ell} + y^{2m} = z^p$

Samuele Anni and Samir Siksek

Let $K$ be a real abelian field of odd class number in which 5 is unramified. Let $S_5$ be the set of places of $K$ above 5. Suppose for every nonempty proper subset $S \subset S_5$ there is a totally positive unit $u \in \mathcal{O}_K$ such that

$$\prod_{q \in S} \text{Norm}_{\mathbb{F}_q/F_5}(u \mod q) \neq 1.$$

We prove that every semistable elliptic curve over $K$ is modular, using a combination of several powerful modularity theorems and class field theory. We deduce that if $K$ is a real abelian field of conductor $n < 100$, with $5 \nmid n$ and $n \neq 29, 87, 89$, then every semistable elliptic curve $E$ over $K$ is modular.

Let $\ell, m, p$ be prime, with $\ell, m \geq 5$ and $p \geq 3$. To a putative nontrivial primitive solution of the generalized Fermat equation $x^{2\ell} + y^{2m} = z^p$ we associate a Frey elliptic curve defined over $\mathbb{Q}(\zeta_p)^+$, and study its mod $\ell$ representation with the help of level lowering and our modularity result. We deduce the nonexistence of nontrivial primitive solutions if $p \leq 11$, or if $p = 13$ and $\ell, m \neq 7$.

1. Introduction

Let $p, q, r \in \mathbb{Z}_{\geq 2}$. The equation

$$x^p + y^q = z^r$$

is known as the generalized Fermat equation (or the Fermat–Catalan equation) with signature $(p, q, r)$. As in Fermat’s last theorem, one is interested in integer solutions $x, y, z$. Such a solution is called nontrivial if $xyz \neq 0$, and primitive if $x, y, z$ are coprime. Let $\chi = p^{-1} + q^{-1} + r^{-1}$. The generalized Fermat conjecture [Darmon and Granville 1995; Darmon 1997], also known as the Tijdeman–Zagier
conjecture and as the Beal conjecture [Beukers 2012], is concerned with the case \( \chi < 1 \). It states that the only nontrivial primitive solutions to (1) with \( \chi < 1 \) are

\[
\begin{align*}
1 + 2^3 &= 3^2, \\
7^3 + 13^2 &= 2^9, \\
3^5 + 11^4 &= 122^2, \\
17^7 + 76271^3 &= 21063928^2, \\
33^8 + 96222^3 &= 30042907^2,
\end{align*}
\]

The conjecture has been established for many signatures \((p, q, r)\), including several infinite families of signatures, starting with Fermat’s last theorem \((p, p, p)\) by Wiles [1995]; \((p, p, 2)\) and \((p, p, 3)\) by Darmon and Merel [1997]; \((2, 4, p)\) by Ellenberg [2004] and Bennett, Ellenberg and Ng [Bennett et al. 2010]; \((2p, 2p, 5)\) by Bennett [2006]; \((2, 6, p)\) by Bennett and Chen [2012]; and other signatures by other researchers. An excellent, exhaustive and up-to-date survey was recently compiled by Bennett, Chen, Dahmen and Yazdani [Bennett et al. 2015a], which also proves the generalized Fermat conjecture for several families of signatures, including \((2p, 4, 3)\).

The main Diophantine result of this paper is the following theorem.

**Theorem 1.1.** Let \( p = 3, 5, 7, 11 \) or 13. Let \( \ell, m \geq 5 \) be primes, and if \( p = 13 \) suppose moreover that \( \ell, m \neq 7 \). Then the only primitive solutions to

\[ x^{2\ell} + y^{2m} = z^p \]

are the trivial ones \((x, y, z) = (\pm 1, 0, 1)\) and \((0, \pm 1, 1)\).

If \( \ell = 2, 3 \) or \( m = 2, 3 \) then (2) has no nontrivial primitive solutions for prime \( p \geq 3 \); this follows from the aforementioned work on Fermat equations of signatures \((2, 4, p)\), \((2, 6, p)\) and \((2p, 4, 3)\).

Our approach is unusual in that it treats several bi-infinite families of signatures. We start with a descent argument (Section 4), inspired by the approach of Bennett [2006] for \( x^{2n} + y^{2n} = z^5 \) and that of Freitas [2015] for \( x^r + y^r = z^p \) with certain small values of \( r \). For \( p = 3 \) the descent argument allows us to quickly obtain a contradiction (Section 5) through results of Bennett and Skinner [2004]. The bulk of the paper is devoted to \( 5 \leq p \leq 13 \). Our descent allows us to construct Frey curves (Sections 6 and 7) attached to (2) that are defined over the real cyclotomic field \( K = \mathbb{Q}(\zeta + \zeta^{-1}) \) where \( \zeta \) is a \( p \)-th root of unity, or, for \( p \equiv 1 \pmod{4} \), defined over the unique subfield \( K' \) of \( K \) of degree \( \frac{1}{4}(p - 1) \). These Frey curves are semistable over \( K \), though not necessarily over \( K' \).

In the remainder of the paper we study the mod \( \ell \) representations of these Frey curves using modularity and level lowering. Several recent papers [Dieulefait and
Modular elliptic curves over real abelian fields and generalized Fermat

Freitas 2013; Freitas and Siksek 2015a; 2015c; Freitas 2015; Bennett et al. 2015b] apply modularity and level lowering over totally real fields to study Diophantine problems. We need to refine many of the ideas in those papers, both because we are dealing with representations over number fields of relatively high degree, and because we are aiming for a “clean” result without any exceptions (the methods are much easier to apply for sufficiently large \( \ell \)). We first establish modularity of the Frey curves by combining a modularity theorem for residually reducible representations due to Skinner and Wiles [1999] with a theorem of Thorne [2016] for residually dihedral representations, and implicitly applying modularity lifting theorems of Kisin [2009] and others for representations with “big image”. We use class field theory to glue together these great modularity theorems and produce our own theorem (proved in Section 2) that applies to our Frey curves, but which we expect to be of independent interest.

**Theorem 1.2.** Let \( K \) be a real abelian number field. Write \( S_5 \) for the prime ideals \( q \) of \( K \) above 5. Suppose

(a) 5 is unramified in \( K \);
(b) the class number of \( K \) is odd;
(c) for each nonempty proper subset \( S \) of \( S_5 \), there is some totally positive unit \( u \) of \( \mathcal{O}_K \) such that

\[
\prod_{q \in S} \text{Norm}_{F_q/F_5}(u \ mod \ q) \neq \overline{1}.
\]

Then every semistable elliptic curve \( E \) over \( K \) is modular.

**Theorem 1.2** allows us to deduce the following corollary (also proved in Section 2).

**Corollary 1.3.** Let \( K \) be a real abelian field of conductor \( n < 100 \) with \( 5 \nmid n \) and \( n \neq 29, 87, 89 \). Let \( E \) be a semistable elliptic curve over \( K \). Then \( E \) is modular.

To apply level lowering theorems to a modular mod \( \ell \) representation, one must first show that this representation is irreducible. Let \( G_K = \text{Gal}(\overline{K}/K) \). The mod \( \ell \) representation that concerns us, denoted \( \overline{\rho}_{E,\ell} : G_K \to \text{GL}_2(\mathbb{F}_\ell) \), is the one attached to the \( \ell \)-torsion of our semistable Frey elliptic curve \( E \) defined over the field \( K = \mathbb{Q}(\zeta + \zeta^{-1}) \) of degree \( \frac{1}{2}(p-1) \). We exploit semistability of our Frey curve to show, with the help of class field theory, that if \( \overline{\rho}_{E,\ell} \) is reducible then \( E \) or some \( \ell \)-isogenous curve possesses nontrivial \( K \)-rational \( \ell \)-torsion. Using famous results on torsion of elliptic curves over number fields of small degree due to Kamienny [1992], Parent [2000; 2003], and Derickx et al. [≥ 2016] and some computations of \( K \)-points on certain modular curves, we prove the required irreducibility result (Section 10).

The final step (Section 11) in the proof of Theorem 1.1 requires computations of certain Hilbert eigenforms over the fields \( K \) together with their eigenvalues at
primes of small norm. For these computations we have made use of the “Hilbert modular forms package” developed by Dembélé, Donnelly, Greenberg and Voight and available within the Magma computer algebra system [Bosma et al. 1997]. For the theory behind this package see [Dembélé and Voight 2013]. For \( p \geq 17 \), the required computations are beyond the capabilities of current software, though the strategy for proving Theorem 1.1 should be applicable to larger \( p \) once these computational limitations are overcome. In fact, at the end of Section 11, we heuristically argue that the larger the value of \( p \) is, the more likely that the argument used to complete the proof of Theorem 1.1 will succeed for that particular \( p \). We content ourselves with proving the following theorem (Section 8).

**Theorem 1.4.** Let \( p \) be an odd prime, and let \( K = \mathbb{Q}(\zeta + \zeta^{-1}) \) for \( \zeta = \exp(2\pi i / p) \). Write \( \mathcal{O}_K \) for the ring of integers in \( K \) and \( \mathfrak{p} \) for the unique prime ideal above \( p \). Suppose that there are no elliptic curves \( E/K \) with full 2-torsion and conductors \( 2\mathcal{O}_K, 2p \). Then there is an ineffective constant \( C_p \) (depending only on \( p \)) such that for all primes \( \ell, m \geq C_p \), the only primitive solutions to (2) are the trivial ones \((x, y, z) = (\pm 1, 0, 1)\) and \((0, \pm 1, 1)\).

If \( p \equiv 1 \pmod{4} \) then let \( K' \) be the unique subfield of \( K \) of degree \( \frac{1}{4}(p-1) \). Let \( \mathfrak{B} \) be the unique prime ideal of \( K' \) above \( p \). Suppose that there are no elliptic curves \( E/K' \) with nontrivial 2-torsion and conductors \( 2\mathfrak{B}, 2\mathfrak{B}^2 \). Then there is an ineffective constant \( C_p \) (depending only on \( p \)) such that for all primes \( \ell, m \geq C_p \), the only primitive solutions to (2) are the trivial ones \((x, y, z) = (\pm 1, 0, 1)\) and \((0, \pm 1, 1)\).

The computations described in this paper were carried out using the computer algebra system Magma [Bosma et al. 1997]. The code and output is available from http://homepages.warwick.ac.uk/∼maseap/progs/diophantine/

2. Proof of Theorem 1.2 and Corollary 1.3

We need a result from class field theory. The following version is proved by Kraus [2007, Appendice A].

**Proposition 2.1.** Let \( K \) be a number field, and \( q \) a rational prime that does not ramify in \( K \). Denote the mod \( q \) cyclotomic character by \( \chi_q : G_K \to \mathbb{F}_q^\times \). Write \( S_q \) for the set of primes \( q \) of \( K \) above \( q \), and let \( S \) be a subset of \( S_q \). Let \( \varphi : G_K \to \mathbb{F}_q^\times \) be a character satisfying:

(a) \( \varphi \) is unramified outside \( S \) and the infinite places of \( K \);

(b) \( \varphi|_{I_q} = \chi_q|_{I_q} \) for all \( q \in S \); here \( I_q \) denotes the inertia subgroup of \( G_K \) at \( q \).

Let \( u \in \mathcal{O}_K \) be a unit that is positive in each real embedding of \( K \). Then

\[
\prod_{q \in S} \text{Norm}_{\mathbb{F}_q / \mathbb{F}_q}(u \mod q) = \bar{1}.
\]
Proof. For the reader’s convenience we give a sketch of Kraus’s elegant argument. Let $L$ be the cyclic field extension of $K$ cut out by the kernel of $\varphi$. Then we may view $\varphi$ as a character $\text{Gal}(L/K) \to \mathbb{F}_q^\times$. Write $M_K$ for the places of $K$. For $\nu \in M_K$, let $\Theta_\nu : K_\nu^* \to \text{Gal}(L/K)$ be the local Artin map. Let $u \in O_K$ be a unit that is positive in each real embedding. We consider the values $\varphi(\Theta_\nu(u)) \in \mathbb{F}_q^\times$ as $\nu$ ranges over $M_K$.

Suppose first that $\nu \in M_K$ is infinite. If $\nu$ is complex then $\Theta_\nu$ is trivial and so certainly $\varphi(\Theta_\nu(u)) = \bar{1}$ in $\mathbb{F}_q^\times$. So suppose $\nu$ is real. As $u$ is positive in $K$, it is a local norm and hence in the kernel of $\Theta_\nu$. Therefore $\varphi(\Theta_\nu(u)) = \bar{1}$.

Suppose next that $\nu \in M_K$ is finite. As $u \in O_\nu^\times$, it follows from local reciprocity that $\Theta_\nu(u)$ belongs to the inertia subgroup $I_\nu \subseteq \text{Gal}(L/K)$. If $\nu \notin S$ then $\varphi(I_\nu) = 1$ by (a) and so $\varphi(\Theta_\nu(u)) = 1$. Thus suppose that $\nu = q \in S$. It follows from (b) that $\varphi(\Theta_q(u)) = \chi_q(\Theta_q(u))$. Through an explicit calculation, Kraus [2007, Appendix A, Proposition 1] shows that $\chi_q(\Theta_q(u)) = \text{Norm}_{\mathbb{F}_q/\mathbb{F}_q}(u \mod q)^{-1}$.

Finally, by global reciprocity, $\prod_{\nu \in M_K} \Theta_\nu(u) = 1$. Applying $\varphi$ to this equality completes the proof.

We also make use of the following theorem of Thorne [2016, Theorem 1.1].

Theorem 2.2 (Thorne). Let $E$ be an elliptic curve over a totally real field $K$. Suppose 5 is not a square in $K$ and that $E$ has no 5-isogenies defined over $K$. Then $E$ is modular.

Thorne deduces this result by combining his beautiful modularity theorem for residually dihedral representations [Thorne 2016, Theorem 1.2], with [Freitas et al. 2015, Theorem 3]. The latter result is essentially a straightforward consequence of the powerful modularity lifting theorems for residual representations with “big image” due to Kisin [2009], Barnet-Lamb et al. [2012; 2013] and Breuil and Diamond [2014].

Finally we need the following modularity theorem for residually reducible representations due to Skinner and Wiles [1999, Theorem A].

Theorem 2.3 (Skinner and Wiles). Let $K$ be a real abelian number field. Let $q$ be an odd prime, and

$$\rho : G_K \to \text{GL}_2(\mathbb{Q}_q)$$

be a continuous, irreducible representation, unramified away from a finite number of places of $K$. Suppose $\bar{\rho}$ is reducible and write $\bar{\rho}^{ss} = \psi_1 \oplus \psi_2$. Suppose further that

(i) the splitting field $K(\psi_1/\psi_2)$ of $\psi_1/\psi_2$ is abelian over $\mathbb{Q}$;
(ii) $(\psi_1/\psi_2)(\tau) = -1$ for each complex conjugation $\tau$;
(iii) $(\psi_1/\psi_2)|_{D_q} \neq 1$ for each $q | q$;
(iv) for all \( q \mid q \),
\[
\rho|_{D_q} \sim \begin{pmatrix} \phi_1^{(q)} \cdot \tilde{\psi}_1 & * \\ 0 & \phi_2^{(q)} \cdot \tilde{\psi}_2 \end{pmatrix}
\]
with \( \phi_2^{(q)} \) factoring through a pro-\( q \) extension of \( K_q \) and \( \phi_2^{(q)}|_{I_q} \) having finite order, and where \( \tilde{\psi}_i \) is a Teichmüller lift of \( \psi_i \);

(v) \( \det(\rho) = \psi \chi_q^{k-1} \), where \( \psi \) is a character of finite order and \( k \geq 2 \) is an integer.

Then the representation \( \rho \) is associated to a Hilbert modular newform.

**Proof of Theorem 1.2.** As 5 is unramified in \( K \), it certainly is not a square in \( K \). If \( E \) has no 5-isogenies defined over \( K \) then the result follows from Thorne’s theorem. We may thus suppose that the mod 5 representation \( \bar{\rho} \) of \( E \) is reducible, and write \( \bar{\rho}^{ss} = \psi_1 \oplus \psi_2 \). We verify hypotheses (i)–(v) in the theorem of Skinner and Wiles (with \( q = 5 \)) to deduce the modularity of \( \rho : G_K \to \text{Aut}(T_5(E)) \cong \text{GL}_2(\mathbb{Z}_5) \), where \( T_5(E) \) is the 5-adic Tate module of \( E \). If \( E \) has good supersingular reduction at some \( q \mid 5 \) then (as \( q \) is unramified) \( \bar{\rho}|_{I_q} \) is irreducible [Serre 1972, Proposition 12], contradicting the reducibility of \( \bar{\rho} \). It follows that \( E \) has good ordinary or multiplicative reduction at all \( q \mid 5 \). In particular, hypothesis (iv) holds with \( \phi_i^{(q)} = 1 \).

Now \( \psi_1 \psi_2 = \det(\rho) = \chi_5 \) so hypothesis (v) holds with \( \psi = 1 \) and \( k = 2 \). Moreover, for each complex conjugation \( \tau \), we have
\[
(\psi_1/\psi_2)(\tau) = \psi_1(\tau)\psi_2(\tau^{-1}) = \psi_1(\tau)\psi_2(\tau) = \chi_5(\tau) = -1,
\]
so (ii) is satisfied. It follows from the fact that \( E \) has good ordinary or multiplicative reduction at all \( q \mid 5 \), that \( (\bar{\rho}|_{I_q})^{ss} = \chi_5|_{I_q} \oplus 1 \) and so \( \psi_1/\psi_2 \) is nontrivial when restricted to \( I_q \) (again as \( q \) is unramified in \( K \)); this proves (iii).

It remains to verify (i). Note that \( \psi_1/\psi_2 = \chi_5/\psi_2^2 \). Hence \( K(\psi_1/\psi_2) \) is contained in the compositum of the fields \( K(\zeta_5) \) and \( K(\psi_2^2) \), and by symmetry also contained in the compositum of the fields \( K(\zeta_5) \) and \( K(\psi_1^2) \). It is sufficient to show that either \( K(\psi_2^2) = K \) or \( K(\psi_1^2) = K \). Note that \( \psi_i^2 : G_K \to \mathbb{F}_5^\times \) are quadratic characters that are unramified at all archimedean places. We will show that one of them is everywhere unramified, and then the desired result follows from the assumption that the class number of \( K \) is odd. First note, by the semistability of \( E \), that \( \psi_1 \) and \( \psi_2 \) are unramified at all finite primes \( p \mid 5 \). Let \( S \) be the subset of \( q \in S_5 \) such that \( \psi_1 \) is unramified at \( q \). By the above, we know that these are precisely the \( q \in S_5 \) such that \( \psi_2|_{I_q} = \chi_5|_{I_q} \). By assumption (c) and Proposition 2.1, we have that either \( S = \emptyset \) or \( S = S_5 \). If \( S = \emptyset \) then \( \psi_2 \) is unramified at all \( q \mid 5 \), and if \( S = S_5 \) then \( \psi_1 \) is unramified at all \( q \mid 5 \). This completes the proof.

**Proof of Corollary 1.3.** Suppose first that \( K = \mathbb{Q}(\zeta_n)^+ \). If \( n \equiv 2 \pmod{4} \) then \( \mathbb{Q}(\zeta_n) = \mathbb{Q}(\zeta_{n/2}) \), so we adopt the usual convention of supposing that \( n \not\equiv 2 \pmod{4} \). We consider values \( n < 100 \) and impose the restriction \( 5 \nmid n \), which ensures that
condition (a) of Theorem 1.2 is satisfied. It is known [Miller 2014] that the class number $h_n^+$ of $K$ is 1 for all $n < 100$. Thus condition (b) is also satisfied. Write $E_n^+$ for the group of units of $K$ and $C_n^+$ for the subgroup of cyclotomic units. A result of Sinnott [1978] asserts that $[E_n^+ : C_n^+] = b \cdot h_n^+$, where $b$ is an explicit constant that happens to be 1 for $n$ with at most 3 distinct prime divisors, and so certainly for all $n$ in our range. It follows that $E_n^+ = C_n^+$ for $n < 100$. Now let $S_5$ be as in the statement of Theorem 1.2. We wrote a simple Magma script which for each $n < 100$ satisfying $5 \nmid n$ and $n \not\equiv 2 \pmod{4}$ writes down a basis for the cyclotomic units $C_n^+$ and deduces a basis for the totally positive units. It then checks, for every nonempty proper subset of $S_5$, if there is an element $u$ of this basis of totally positive units that satisfies (3). We found this to be the case for all $n$ under consideration except $n = 29, 87$ and 89. The corollary follows from Theorem 1.2 for $K = \mathbb{Q}(\zeta_n)^+$ with $n$ as in the statement of the corollary.

Now let $K$ be a real abelian field with conductor $n$ as in the statement of the corollary. Then $K \subseteq \mathbb{Q}(\zeta_n)^+$. As $\mathbb{Q}(\zeta_n)^+/K$ is cyclic, modularity of an elliptic curve $E/K$ follows, by Langlands’ cyclic base change theorem [Langlands 1980], from modularity of $E$ over $\mathbb{Q}(\zeta_n)^+$, completing the proof of the corollary. □

3. Cyclotomic preliminaries

Throughout $p$ will be an odd prime. Let $\zeta$ be a primitive $p$-th root of unity, and $K = \mathbb{Q}(\zeta + \zeta^{-1})$ the maximal real subfield of $\mathbb{Q}(\zeta)$. We write

$$\theta_j = \zeta^j + \zeta^{-j} \in K, \quad j = 1, \ldots, \frac{1}{2}(p - 1).$$

Let $O_K$ be the ring of integers of $K$. Let $p$ be the unique prime ideal of $K$ above $p$. Then $pO_K = p^{(p-1)/2}$.

**Lemma 3.1.** For $j = 1, \ldots, \frac{1}{2}(p - 1)$, we have

$$\theta_j \in O_K^\times, \quad \theta_j + 2 \in O_K^\times, \quad (\theta_j - 2)O_K = p.$$

Moreover, $(\theta_j - \theta_k)O_K = p$ for $1 \leq j < k \leq \frac{1}{2}(p - 1)$.

**Proof.** Observe that $\theta_j = (\zeta^{2j} - \zeta^{-2j})/(\zeta^j - \zeta^{-j})$ and thus belongs to the group of cyclotomic units. Given $j$, let $j \equiv 2r \pmod{p}$. Then $\theta_j + 2 = \theta_r^2 \in O_K^\times$.

For now, let $L = \mathbb{Q}(\zeta)$. Let $\mathfrak{p}$ be the prime of $O_L$ above $p$. Then $pO_L = \mathfrak{p}^2$. As is well-known, $\mathfrak{p} = (1 - \zeta^u)O_L$ for $u = 1, 2, \ldots, p - 1$. Note that $\theta_j - 2 = (\zeta^r - \zeta^{-r})^2$, with $j \equiv 2r \pmod{p}$, from which we deduce that $(\theta_j - 2)O_L = \mathfrak{p}^2 = pO_L$, hence $(\theta_j - 2)O_K = p$.

For the final part, $j \not\equiv \pm k \pmod{p}$. Thus there exist $u, v \not\equiv 0 \pmod{p}$ such that $u + v \equiv j, \quad u - v \equiv k \pmod{p}$.
Then
\[(\zeta^u - \zeta^{-u})(\zeta^v - \zeta^{-v}) = \theta_j - \theta_k,\]
and so \((\theta_j - \theta_k)\mathcal{O}_L = \mathfrak{P}^2 = p\mathcal{O}_L\). This completes the proof. \(\square\)

4. The descent

Now let \(\ell, m \geq 5\) be prime, and let \((x, y, z)\) be a nontrivial, primitive solution to (2). If \(\ell = p\), then (2) can be rewritten as \(z^p + (-x^2)^p = (y^m)^2\). Darmon and Merel [1997] have shown that the only primitive solutions to the generalized Fermat equation (1) with signature \((p, p, 2)\) are the trivial ones, giving us a contradiction. We shall henceforth suppose that \(\ell \neq p\) and \(m \neq p\).

Clearly \(z\) is odd. By swapping in (2) the terms \(x^\ell\) and \(y^m\) if necessary, we may suppose that \(2 \mid x\). We factor the left-hand side over \(\mathbb{Z}[i]\). It follows from our assumptions that the two factors \(x^\ell + y^mi\) and \(x^\ell - y^mi\) are coprime. There exist coprime rational integers \(a, b\) such that
\[x^\ell + y^mi = (a + bi)^p, \quad z = a^2 + b^2.\]

Then
\[x^\ell = \frac{1}{2}((a + bi)^p + (a - bi)^p)\]
\[= a \cdot \prod_{j=1}^{(p-1)/2} ((a + bi) + (a - bi)\zeta^j)\]
\[= a \cdot \prod_{j=1}^{(p-1)/2} ((a + bi) + (a - bi)\zeta^j) \cdot ((a + bi) + (a - bi)\zeta^{-j}).\]

In the last step we have paired up the complex conjugate factors. Multiplying out these pairs we obtain a factorization of \(x^\ell\) over \(\mathcal{O}_K\):
\[x^\ell = a \cdot \prod_{j=1}^{(p-1)/2} ((\theta_j + 2)a^2 + (\theta_j - 2)b^2).\] (4)

To ease notation, write
\[\beta_j = (\theta_j + 2)a^2 + (\theta_j - 2)b^2, \quad j = 1, \ldots, \frac{1}{2}(p - 1).\] (5)

Lemma 4.1. Write \(n = \nu_2(x) \geq 1\).

(i) If \(p \nmid x\) then
\[a = 2^\ell n \alpha^\ell, \quad \beta_j \mathcal{O}_K = b_j^\ell,\]
where \(\alpha\) is a rational integer and \(\alpha \mathcal{O}_K, b_1, \ldots, b_{(p-1)/2}\) are pairwise coprime ideals of \(\mathcal{O}_K\), all of which are coprime to \(2p\).
(ii) If \( p \mid x \) then
\[
a = 2^{\ell n} p^{\kappa \ell - 1} \alpha^\ell, \quad \beta_j \mathcal{O}_K = p \cdot b_j^\ell,
\]
where \( \kappa = \nu_p(x) \geq 1 \), \( \alpha \) is a rational integer and \( \alpha \cdot \mathcal{O}_K, b_1, \ldots, b_{(p-1)/2} \) are pairwise coprime ideals of \( \mathcal{O}_K \), all of which are coprime to \( 2p \).

**Proof.** As \( z = a^2 + b^2 \) is odd, exactly one of \( a, b \) is even. Thus the \( \beta_j \) are coprime to \( 2 \mathcal{O}_K \). We see from (4) that \( 2^{\ell n} \parallel a \), and hence that \( b \) is odd.

As \( a, b \) are coprime, it is clear that the greatest common divisor of \( a \mathcal{O}_K \) and \( \beta_j \mathcal{O}_K \) divides \( \theta_j - 2 \mathcal{O}_K = \mathfrak{p} \). Moreover, for \( k \neq j \), the greatest common divisor of \( \beta_j \mathcal{O}_K \) and \( \beta_k \mathcal{O}_K \) divides
\[
((\theta_j + 2)(\theta_k - 2) - (\theta_k + 2)(\theta_j - 2)) \mathcal{O}_K = 4(\theta_k - \theta_j) \mathcal{O}_K = 4 \mathfrak{p}.
\]
However, \( \beta_j \) is odd, and so the greatest common divisor of \( \beta_j \mathcal{O}_K \) and \( \beta_k \mathcal{O}_K \) divides \( \mathfrak{p} \). Now (i) follows immediately from (4). So suppose \( p \mid x \). For (ii) we have to check that \( p \parallel \beta_j \). However, since \( (\theta_j - 2) \mathcal{O}_K = \mathfrak{p} \) and \( \theta_j + 2 \in \mathcal{O}_K^\times \), reducing (4) modulo \( p \) shows that \( a^p \equiv 0 \pmod{p} \), and hence that \( p \mid a \). Since \( a, b \) are coprime, it follows that \( \nu_p(\beta_j) = 1 \). Now, from (4),
\[
\frac{1}{2}(p - 1) \nu_p(a) = \nu_p(a) = \ell \nu_p(x) - \sum_{j=1}^{(p-1)/2} \nu_p(\beta_j) = \frac{1}{2}(p - 1)(\kappa \ell - 1),
\]
giving the desired exponent of \( p \) in the factorization of \( a \). \( \square \)

5. **Proof of Theorem 1.1 for \( p = 3 \)**

Suppose \( p = 3 \). Then \( K = \mathbb{Q} \) and \( \theta := \theta_1 = -1 \). We treat first the case \( 3 \nmid x \). By Lemma 4.1,
\[
a = 2^{\ell n} \alpha^\ell, \quad a^2 - 3b^2 = \gamma^\ell
\]
for some coprime odd rational integers \( \alpha \) and \( \gamma \). We obtain the equation
\[
2^{2\ell n} \alpha^{2\ell} - \gamma^\ell = 3b^2.
\]
Bennett and Skinner [2004, Theorem 1] show that the equation \( x^n + y^n = 3z^2 \) has no solutions in coprime integers \( x, y, z \) for \( n \geq 4 \), giving us a contradiction.

We now treat \( 3 \mid x \). By Lemma 4.1,
\[
a = 2^{\ell n} 3^{\kappa \ell - 1} \alpha^\ell, \quad a^2 - 3b^2 = 3\gamma^\ell
\]
for coprime rational integers \( \alpha, \gamma \) that are also coprime to 6. Thus
\[
2^{2\ell n} 3^{2\kappa \ell - 3} \alpha^{2\ell} - \gamma^\ell = b^2.
\]
Using the recipes of Bennett and Skinner [2004, Sections 2, 3], we can attach a Frey curve to such a triple \((\alpha, \gamma, b)\) whose mod \(\ell\) representation arises from a classical newform of weight 2 and level 6. As there are no such newforms our contradiction is complete.

6. The Frey curve

We shall henceforth suppose \(p \geq 5\). From now on, fix \(1 \leq j, k \leq \frac{1}{2} (p - 1)\) with \(j \neq k\). The expressions \(\beta_j, \beta_k\) are given by (5). For each such choice of \((j, k)\) we shall construct a Frey curve. The idea is that the three expressions \(a^2, \beta_j, \beta_k\) are roughly \(\ell\)-th powers (Lemma 4.1). Moreover they are linear combinations of \(a^2\) and \(b^2\), and hence must be linearly dependent. Writing down this linear relation gives a Fermat equation (with coefficients) of signature \((\ell, \ell, \ell)\). As in the work of Hellegouarch, Frey, Serre, Ribet, Kraus and many others, one can associate to such an equation a Frey elliptic curve whose mod \(\ell\) representation has very little ramification. In what follows we take care to scale the expressions \(a^2, \beta_j, \beta_k\) appropriately so that the Frey curve is semistable.

Case I: \(p \nmid x\). Let

\[
\begin{align*}
    u &= \beta_j, \\
    v &= -\frac{(\theta_j - 2)}{(\theta_k - 2)} \beta_k, \\
    w &= \frac{4(\theta_j - \theta_k)}{(\theta_k - 2)} \cdot a^2.
\end{align*}
\]

Then \(u + v + w = 0\). Moreover, by Lemmas 3.1 and 4.1,

\[
\begin{align*}
    uO_K &= b_j^\ell, \\
    vO_K &= b_k^\ell, \\
    wO_K &= 2^{2\ell n + 2} \cdot \alpha^2 \cdot b_j \cdot b_k.
\end{align*}
\]

We let the Frey curve be

\[
E = E_{j,k} : Y^2 = X(X - u)(X + v).
\]

For a nonzero ideal \(a\), we define its radical, denoted by \(\text{Rad}(a)\), to be the product of the distinct prime ideal factors of \(a\).

Lemma 6.1. Suppose \(p \nmid x\). Let \(E\) be the Frey curve (7) where \(u, v, w\) are given by (6). The curve \(E\) is semistable, with multiplicative reduction at all primes above 2 and good reduction at \(p\). It has minimal discriminant and conductor

\[
\begin{align*}
    \Delta_{E/K} &= 2^{4\ell n - 4} \alpha^{4\ell} b_j^{2\ell} b_k^{2\ell}, \\
    N_{E/K} &= 2 \cdot \text{Rad}(\alpha b_j b_k).
\end{align*}
\]

Proof. The invariants \(c_4, c_6, \Delta, j(E)\) have their usual meanings and are given by

\[
\begin{align*}
    c_4 &= 16(u^2 - vw) = 16(v^2 - wu) = 16(w^2 - uv), \\
    c_6 &= -32(u - v)(v - w)(w - u), \\
    \Delta &= 16u^2v^2w^2, \\
    j(E) &= c_4^3/\Delta.
\end{align*}
\]

By Lemma 4.1, we have that \(\alpha O_K, b_j, b_k\) are pairwise coprime, and all coprime to \(2p\). In particular \(p \nmid \Delta\) and so \(E\) has good reduction at \(p\). Moreover, \(c_4\) and \(\Delta\) are
coprime away from 2. Hence the model in (7) is already semistable away from 2. Recall that \( 2^\ell \mid a \) and \( 2 \nmid b \). Thus

\[
u \equiv (\theta_j - 2)b^2 \mod {2^\ell}, \quad v \equiv -((\theta_j - 2)b^2) \mod {2^\ell}, \quad w \equiv 0 \mod {2^{2\ell+2}}.
\]

It is clear that \( v_q(j) < 0 \) for all \( q \mid 2 \). Thus \( E \) has potentially multiplicative reduction at all \( q \mid 2 \). Write \( \gamma = -c_4/c_6 \). To show that \( E \) has multiplicative reduction at \( q \) it is enough to show that \( K_q(\sqrt{\gamma})/K_q \) is an unramified extension [Silverman 1994, Exercise V.5.11]. However,

\[
\frac{1}{16}c_4 = (u^2 - vw) = (\theta_j - 2)^2 \cdot b^4 \mod {2^\ell},
\]

which shows that \( c_4 \) is a square in \( K_q \). Moreover,

\[
-\frac{1}{16}c_6 = 2(u - v)(w - u) = 4 \cdot (\theta_j - 2)^3 \cdot b^6 \mod {2^{2\ell+1}}.
\]

Thus \( K_q(\sqrt{\gamma}) = K_q(\sqrt[\ell]{\theta_j - 2}) \). As before, letting \( r \) satisfy \( 2r \equiv j \mod p \), we have \( \theta_j - 2 = (\zeta^r - \zeta^{-r})^2 \) and so \( K_q(\sqrt{\gamma}) \) is contained in the unramified extension \( K_q(\zeta) \). Hence \( E \) has multiplicative reduction at \( q \mid 2 \).

Finally 2 is unramified in \( K \), and so \( v_q(c_4) = v_2(16) = 4 \). It follows that \( D_{E/K} = (\Delta/2^{12}) \cdot \mathcal{O}_K \), as required. \( \square \)

**Case II:** \( p \mid x \). Let

\[
u = -\frac{\beta_k}{(\theta_k - 2)}, \quad u = \frac{\beta_j}{(\theta_j - 2)}, \quad w = \frac{4(\theta_j - \theta_k)}{(\theta_j - 2)(\theta_k - 2)} \cdot a^2.
\]

Then, from Lemmas 3.1 and 4.1,

\[
u \mathcal{O}_K = b_j^\ell, \quad u \mathcal{O}_K = b_j^\ell, \quad w \mathcal{O}_K = 2^{2\ell n + 2} \cdot p^\delta \cdot \alpha^{2\ell} \mathcal{O}_K,
\]

where

\[
\delta = (\kappa \ell - 1)(p - 1) - 1.
\]

Again \( u + v + w = 0 \) and the Frey curve is given by (7).

**Lemma 6.2.** Suppose \( p \mid x \). Let \( E \) be the Frey curve (7) where \( u, v, w \) are given by (9). The curve \( E \) is semistable, with multiplicative reduction at \( p \) and at all primes above 2. It has minimal discriminant and conductor

\[
D_{E/K} = 2^{4\ell n - 4}p^{2\ell} \alpha^{4\ell} b_j^\ell b_k^\ell, \quad \mathcal{N}_{E/K} = 2p \cdot \text{Rad}(\alpha b_j b_k).
\]

**Proof.** The proof is an easy modification of the proof of Lemma 6.1. \( \square \)

**7. A closer look at the Frey curve for \( p \equiv 1 \mod 4 \)**

In this section we suppose that \( p \equiv 1 \mod 4 \). The Galois group of \( K = \mathbb{Q}(\zeta + \zeta^{-1}) \) is cyclic of order \( \frac{1}{2}(p - 1) \). Thus the field \( K = \mathbb{Q}(\zeta + \zeta^{-1}) \) has a unique involution \( \tau \in \text{Gal}(K/\mathbb{Q}) \), and we let \( K' \) be the subfield of degree \( \frac{1}{4}(p - 1) \) that is fixed by
this involution. In the previous section we let \(1 \leq j, k \leq \frac{1}{2}(p - 1)\) with \(j \neq k\). In this section we shall impose the further condition that \(\tau(\theta_j) = \theta_k\). Now a glance at the definition (7) of the Frey curve \(E\) and the formulae (9) for \(u\) and \(v\) in the case \(p \mid x\) shows that the curve \(E\) is in fact defined over \(K'\). This is not true in the case \(p \nmid x\), but we can take a twist of the Frey curve so that it is defined over \(K'\).

**Case I:** \(p \nmid x\). Let

\[
\begin{align*}
    u' &= (\theta_k - 2)\beta_j, \\
    v' &= - (\theta_j - 2)\beta_k, \\
    w' &= 4(\theta_j - \theta_k) \cdot a^2,
\end{align*}
\]

and let

\[
E' : Y^2 = X(X - u')(X + v').
\]

Clearly the coefficients of \(E'\) are invariant under \(\tau\), and so \(E'\) is defined over \(K'\). Moreover, \(E'/K\) is the quadratic twist of \(E/K\) by \((\theta_k - 2)\). Let \(B\) be the unique prime of \(K'\) above \(p\). Let

\[
b_{j,k} = \text{Norm}_{K/K'}(b_j) = \text{Norm}_{K/K'}(b_k).
\]

It follows from Lemma 4.1 that the \(O_{K'}\)-ideal \(b_{j,k}\) is coprime to \(\alpha\) and to \(2p\). An easy calculation leads us to the following lemma.

**Lemma 7.1.** Suppose \(p \nmid x\). Let \(E'/K'\) be the above Frey elliptic curve. Then \(E'\) is semistable away from \(B\), with minimal discriminant and conductor

\[
D_{E'/K'} = 2^{4\ell n - 4} B^3 \alpha^4 b_{j,k}^{2\ell}, \\
N_{E'/K'} = 2 \cdot B^2 \cdot \text{Rad}(\alpha b_{j,k}).
\]

**Case II:** \(p \mid x\). Another straightforward computation yields the following lemma.

**Lemma 7.2.** Suppose \(p \mid x\). Let \(E' = E\) be the Frey curve in Lemma 6.2. Then \(E'\) is defined over \(K'\). The curve \(E'/K'\) is semistable with minimal discriminant and conductor

\[
D_{E'/K'} = 2^{4\ell n - 4} B^3 \alpha^4 b_{j,k}^{2\ell}, \\
N_{E'/K'} = 2 \cdot B \cdot \text{Rad}(\alpha b_{j,k}),
\]

where \(\delta\) is given by (10).

**Remark.** Clearly \(E\) has full 2-torsion over \(K\). The curve \(E'\) has a point of order 2 over \(K'\), but not necessarily full 2-torsion.

### 8. Proof of Theorem 1.4

**Lemma 8.1.** Let \(p\) be an odd prime. There is an ineffective constant \(C_p^{(1)}\) depending on \(p\) such that for odd primes \(\ell, m \geq C_p^{(1)}\) and any nontrivial primitive solution \((x, y, z)\) of (2), the Frey curve \(E/K\) as in Section 6 is modular. If \(p \equiv 1 \pmod{4}\) then under the same assumptions, the Frey curve \(E'/K'\) as in Section 7 is modular.
Proof. Freitas et al. [2015] show that for any totally real field $K$ there are at most finitely many nonmodular $j$-invariants. Let $q$ be a prime of $K$ above 2. By Lemmas 6.1 and 6.2, we have $v_q(j(E)) = -(4\ell n - 4)$ with $n \geq 1$. Thus for $\ell, m$ sufficiently large we have $v_q(j(E)) < v_q(j_i)$ for $i = 1, \ldots, r$, completing the proof. □

Remarks. • The argument in [Freitas et al. 2015] relies on Faltings’ theorem (finiteness of the number of rational points on a curve of genus $\geq 2$) to deduce finiteness of the list of possibly nonmodular $j$-invariants. It is for this reason that the constant $C_p^{(1)}$ (and hence the constant $C_p$ in Theorem 1.4) is ineffective.

• In the above argument, it seems that it is enough to suppose that $\ell$ is sufficiently large without an assumption on $m$. However, in Section 4 we swapped the terms $x^{2\ell}$ and $y^{2m}$ in (2) if needed to ensure that $x$ is even. Thus in the above argument we need to suppose that both $\ell$ and $m$ are sufficiently large.

We shall make use of the following result due to Freitas and Siksek [2015b, Theorem 2]. It is a variant of results proved by Kraus [2007] and by David [2012]. All these build on the celebrated uniform boundedness theorem of Merel [1996].

Theorem 8.2. Let $K$ be a totally real field. There is an effectively computable constant $C_K$ such that for a prime $\ell > C_K$, and for an elliptic curve $E/K$ semistable at all $\lambda | \ell$, the mod $\ell$ representation $\bar{\rho}_{E,\ell} : G_K \rightarrow \text{GL}_2(\mathbb{F}_\ell)$ is irreducible.

In [Freitas and Siksek 2015b, Theorem 2] it is assumed that $K$ is Galois as well as totally real. Theorem 8.2 follows immediately on replacing $K$ with its Galois closure.

Lemma 8.3. Let $E/K$ be the Frey curve given in Section 6. Suppose $\bar{\rho}_{E,\ell}$ is irreducible and $E$ is modular. Then $\bar{\rho}_{E,\ell} \sim \bar{\rho}_{f,\lambda}$ for some Hilbert cuspidal eigenform $f$ over $K$ of parallel weight 2 that is new at level $N', where$

$$N'_\ell = \begin{cases} 2\mathcal{O}_K & \text{if } p \nmid x, \\ 2p & \text{if } p \mid x. \end{cases}$$

Here $\lambda | \ell$ is a prime of $\mathbb{Q}_f$, the field generated over $\mathbb{Q}$ by the eigenvalues of $f$.

If $p \equiv 1 \pmod{4}$, let $E'/K'$ be the Frey curve given in Section 7. Suppose $\bar{\rho}_{E',\ell}$ is irreducible and $E$ is modular. Then $\bar{\rho}_{E',\ell} \sim \bar{\rho}_{f,\lambda}$ for some Hilbert cuspidal eigenform $f$ over $K$ of parallel weight 2 that is new at level $N'_\ell$, where

$$N'_\ell = \begin{cases} 2\mathfrak{B}^2 & \text{if } p \nmid x, \\ 2\mathfrak{B} & \text{if } p \mid x. \end{cases}$$

Proof. This is immediate from Lemmas 6.1, 6.2, 7.1 and 7.2, and a standard level lowering recipe derived in [Freitas and Siksek 2015a, Section 2.3] from the work of Jarvis, Fujiwara and Rajaei. Alternatively, one could use modern modularity lifting
theorems which integrate level lowering with modularity lifting, as for example in [Breuil and Diamond 2014].

Proof of Theorem 1.4. Let $K = \mathbb{Q}(\zeta + \zeta^{-1})$ and $E$ be the Frey curve constructed in Section 6. Let $C_p^{(1)}$ be the constant in Lemma 8.1, and $C_p^{(2)} = C_K$ be the constant in Theorem 8.2. Let $C_p = \max(C_p^{(1)}, C_p^{(2)})$. Suppose that $\ell, m \geq C_p$. Then $\tilde{\rho}_{E,\ell}$ is irreducible and modular, and it follows from Lemma 8.3 that $\tilde{\rho}_{E,\ell} \sim \tilde{\rho}_{f,\lambda}$ for some Hilbert eigenform over $K$ of parallel weight 2 that is new at level $N\ell$, where $N\ell = 2O_K$ or $2p$. Now a standard argument (see [Bennett and Skinner 2004, Section 4], [Kraus 1997, Section 3] or [Siksek 2012, Section 9]) shows that, after enlarging $C_p$ by an effective amount, we may suppose that the field of eigenvalues of $f$ is $\mathbb{Q}$. Observe that the level $N\ell$ is nonsquarefull (meaning there is a prime $q$ at which the level has valuation 1). As the level is nonsquarefull and the field of eigenvalues is $\mathbb{Q}$, the eigenform $f$ is known to correspond to some elliptic curve $E_1/K$ [Blasius 2004], and $\tilde{\rho}_{E,\ell} \sim \tilde{\rho}_{E_1,\ell}$. Finally, and again by standard arguments (see one of the references a few lines above), we may enlarge $C_p$ by an effective constant so that the isomorphism $\tilde{\rho}_{E,\ell} \sim \tilde{\rho}_{E_1,\ell}$ forces $E_1$ to either have full 2-torsion, or to be isogenous to an elliptic curve $E_2/K$ that has full 2-torsion. This contradicts the hypothesis of Theorem 1.4 that there are no such elliptic curves with conductor $2O_K$, $2p$, and completes the proof of the first part of the theorem. The proof of the second part is similar, and makes use of the Frey curve $E'/K'$.

9. Modularity of the Frey curve for $5 \leq p \leq 13$

Lemma 9.1. If $p = 5, 7, 11$ or 13 then the Frey curve $E/K$ in Section 6 is modular.

If $p = 5, 13$ then the Frey curve $E'/K'$ in Section 7 is modular.

Proof. Recall that $E$ is defined over $K = \mathbb{Q}(\zeta + \zeta^{-1})$, where $\zeta$ is a primitive $p$-th root of unity. If $p = 5$ then $K = \mathbb{Q}(\sqrt{5})$, and modularity of elliptic curves over real quadratic fields was recently established by Freitas et al. [2015].

For $p = 7, 11, 13$, the prime 5 is unramified in $K$, the class number of $K$ is 1 and condition (c) of Theorem 1.2 is easily verified. Thus $E$ is modular.

For $p = 13$, the curves $E$ and $E'$ are at worst quadratic twists over $K$, and $K/K'$ is quadratic. The modularity of $E'/K'$ follows from the modularity of $E/K$ and the cyclic base change theorem of [Langlands 1980]. For $p = 5$ we could use the same argument, or more simply note that $K' = \mathbb{Q}$, and conclude by the modularity theorem over the rationals [Breuil et al. 2001].

10. Irreducibility of $\tilde{\rho}_{E,\ell}$ for $5 \leq p \leq 13$

We let $E$ be the Frey curve as in Section 6, and $p = 5, 7, 11, 13$. To apply Lemma 8.3 we need to prove the irreducibility of $\tilde{\rho}_{E,\ell}$ for $\ell \geq 5$; equivalently, we need to show
that \( E \) does not have an \( \ell \)-isogeny for \( \ell \geq 5 \). Alas, there is not yet a uniform boundedness theorem for isogenies. The papers [Kraus 2007; David 2012; Freitas and Siksek 2015b] do give effective bounds \( C_K \) such that for \( \ell > C_K \) the representation \( \bar{\rho}_{E,\ell} \) is irreducible, however these bounds are too large for our present purpose. We refine the arguments in those papers, making use of the fact that the curve \( E \) is semistable, and the number fields \( K = \mathbb{Q}(\zeta + \zeta^{-1}) \) all have narrow class number 1. Before doing this, we relate, for \( p = 5 \) and 13, the representations \( \bar{\rho}_{E,\ell} \) and \( \bar{\rho}_{E',\ell} \), where \( E' \) is the Frey curve in Section 7.

**Lemma 10.1.** Suppose \( p = 5 \) or 13. Let \( \tau \) be the unique involution of \( K \), and \( K' \) the subfield fixed by it. Let \( j \) and \( k \) satisfy \( \tau(\theta_j) = \theta_k \). Let \( E/K \) be the Frey elliptic curve in Section 6 and \( E'/K' \) the Frey curve in Section 7, associated to this pair \((j,k)\). Then \( \bar{\rho}_{E,\ell} \) is irreducible as a representation of \( G_K \) if and only if \( \bar{\rho}_{E',\ell} \) is irreducible as a representation of \( G_{K'} \).

**Proof.** Note that \( K/K' \) is a quadratic extension and \( E/K \) is a quadratic twist of \( E'/K \). Thus \( \bar{\rho}_{E,\ell} \) is a twist of \( \bar{\rho}_{E',\ell}|_{G_K} \) by a quadratic character. If \( \bar{\rho}_{E',\ell} \) is reducible as a representation of \( G_{K'} \) then certainly \( \bar{\rho}_{E,\ell} \) is reducible as a representation of \( G_K \).

Conversely, suppose \( \bar{\rho}_{E',\ell}(G_{K'}) \) is irreducible. We would like to show that \( \bar{\rho}_{E,\ell}(G_K) \) is irreducible. It is enough to show that \( \bar{\rho}_{E',\ell}(G_K) \) is irreducible. Let \( q \mid 2 \) be a prime of \( K' \). Then \( u_q(j(E')) = 4 - 4\ell n \), which is negative but not divisible by \( \ell \). Thus \( \bar{\rho}_{E',\ell}(G_{K'}) \) contains an element of order \( \ell \) [Silverman 1994, Proposition V.6.1]. By Dickson’s classification [Swinnerton-Dyer 1973] of subgroups of \( \text{GL}_2(\mathbb{F}_\ell) \) we see that \( \bar{\rho}_{E',\ell}(G_{K'}) \) must contain \( \text{SL}_2(\mathbb{F}_\ell) \). The latter is a simple group, and must therefore be contained in \( \bar{\rho}_{E',\ell}(G_K) \). This completes the proof. \( \square \)

**Lemma 10.2.** Suppose \( \bar{\rho}_{E,\ell} \) is reducible. Then \( E/K \) either has nontrivial \( \ell \)-torsion, or is \( \ell \)-isogenous to an elliptic curve defined over \( K \) that has nontrivial \( \ell \)-torsion.

**Proof.** Suppose \( \bar{\rho}_{E,\ell} \) is reducible, and write

\[
\bar{\rho}_{E,\ell} \sim \begin{pmatrix} \psi_1 & * \\ 0 & \psi_2 \end{pmatrix}.
\]

We show that either \( \psi_1 \) or \( \psi_2 \) is trivial. It follows in the former case that \( E \) has nontrivial \( \ell \)-torsion, and in the latter case that the \( K \)-isogenous curve \( E/\ker(\psi_1) \) has nontrivial \( \ell \)-torsion.

As \( K \) has narrow class number 1 for \( p = 5, 7, 11, 13 \), it is sufficient to show that one of \( \psi_1, \psi_2 \) is unramified at all finite places. As \( E \) is semistable, the characters \( \psi_1 \) and \( \psi_2 \) are unramified away from \( \ell \) and the infinite places. Let \( S_\ell \) be the set of primes \( \lambda \mid \ell \) of \( K \). Let \( S \subset S_\ell \) for the set of \( \lambda \in S_\ell \) such that \( \psi_1 \) is ramified at \( \lambda \). Then \( \psi_2 \) is ramified exactly at the primes \( S \setminus S_\ell \) (see proof of Theorem 1.2).
Moreover, $\psi_1|_{I_{\lambda}} = \chi_\ell|_{I_{\lambda}}$ for all $\lambda \in S$, and $\psi_2|_{I_{\lambda}} = \chi_\ell|_{I_{\lambda}}$ for all $\lambda \in S_\ell \setminus S$. It is enough to show that either $S$ is empty or $S_\ell \setminus S$ is empty.

Suppose $S$ is a nonempty proper subset of $S_\ell$. Fix $\lambda \in S$ and let

$$D = D_\lambda \subset G = \text{Gal}(K/Q)$$

be the decomposition group of $\lambda$; by definition $\lambda^\sigma = \lambda$ for all $\sigma \in D_\lambda$. As $K/Q$ is abelian and Galois, $D_{\lambda'} = D$ for all $\lambda' \in S_\ell$, and $G/D$ acts transitively and freely on $S_\ell$. Fix a set $T$ of coset representatives for $G/D$. Then there is a subset $T' \subset T$ such that

$$S = \{\lambda^{-1}: \tau \in T\}, \quad S_\ell \setminus S = \{\lambda^{-1}: \tau \in T \setminus T'\}.$$

As $S$ is a nonempty proper subset of $S_\ell$, we have that $T'$ is a nonempty proper subset of $T$. Now, by Proposition 2.1, for any totally positive unit $u$ of $O_K$,

$$\prod_{\tau \in T'} \text{Norm}_{\mathbb{F}_\ell/\mathbb{F}_\ell}(u + \lambda^{-1}) = 1.$$

But

$$\text{Norm}_{\mathbb{F}_\ell/\mathbb{F}_\ell}(u + \lambda^{-1}) = \prod_{\sigma \in D} (u + \lambda^{-1})^\sigma$$

$$= \prod_{\sigma \in D} (u^\sigma + \lambda^{-1})$$

$$= \left(\prod_{\sigma \in D} (u^\sigma + \lambda)\right)^{-1}$$

$$= \prod_{\sigma \in D} (u^\sigma + \lambda)$$

(as this expression belongs to $\mathbb{F}_\ell$).

Let

$$B_{T',D}(u) = \text{Norm}_{K/Q}\left(\left(\prod_{\tau \in T',\sigma \in D} u^{\sigma \tau}\right) - 1\right).$$

It follows that $\ell | B_{T',D}(u)$. Now let $u_1, \ldots, u_d$ be a system of totally positive units. Then $\ell$ divides

$$B_{T',D}(u_1, \ldots, u_d) = \gcd(B_{T',D}(u_1), \ldots, B_{T',D}(u_d)).$$

To sum up, if the lemma is false for $\ell$, then there is some subgroup $D$ of $G$ and some nonempty proper subset $T'$ of $G/D$ such that $\ell$ divides $B_{T',D}(u_1, \ldots, u_d)$.

The proof of the lemma is completed by a computation that we now describe. For each of $p = 5, 7, 11, 13$ we fix a basis $u_1, \ldots, u_d$ for the system of totally positive units of $O_K$. We run through the subgroups $D$ of $G = \text{Gal}(K/Q)$. For each subgroup $D$ we fix a set of coset representatives $T$, and run through the nonempty proper subsets $T'$ of $T$, computing $B_{T',D}(u_1, \ldots, u_d)$. We found that for $p = 5, 7$
the possible values for $B_{T,D}(u_1, \ldots, u_d)$ are all 1; for $p = 11$ they are 1 and 23; and for $p = 13$ they are 1, 5$^2$ and 3$^5$. Thus the proof is complete for $p = 5, 7$ and it remains to deal with $(p, \ell) = (11, 23), (13, 5)$. For each of these possibilities we run through the nonempty proper $S \subset S_i$ and check that there is some totally positive unit $u$ such that $\prod_{\lambda \in S} \text{Norm}(u + \lambda) \neq \bar{1}$. This completes the proof. \hfill \Box

Suppose $\bar{\rho}_{E,\ell}$ is reducible. It follows from Lemma 10.2 that there is an elliptic curve $E_1/K$ (which is either $E$ or $\ell$-isogenous to $E$) such that $E_1(K)$ has a subgroup isomorphic to $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\ell\mathbb{Z}$. Such an elliptic curve is isogenous$^1$ to an elliptic curve $E_2/K$ with a $K$-rational cyclic subgroup isomorphic to $\mathbb{Z}/4\ell\mathbb{Z}$. Thus we obtain a noncuspidal $K$-point on the curves

$$X_0(\ell), \ X_1(\ell), \ X_0(2\ell), \ X_1(2\ell), \ X_0(4\ell), \ X_1(2, 2\ell).$$

To achieve a contradiction it is enough to show that there are no noncuspidal $K$-points on one of these curves. For small values of $\ell$, we find Magma’s “small modular curves package”, as well as Magma’s functionality for computing Mordell–Weil groups of elliptic curves over number fields, invaluable. Four of the modular curves of interest to us happen to be elliptic curves. The aforementioned Magma package gives the following models:

$$X_0(20) : y^2 = x^3 + x^2 + 4x + 4 \quad \text{(Cremona label 20a1)}, \quad (12)$$

$$X_0(14) : y^2 + xy + y = x^3 + 4x - 6 \quad \text{(Cremona label 14a1)}, \quad (13)$$

$$X_0(11) : y^2 + y = x^3 - x^2 - 10x - 20 \quad \text{(Cremona label 11a1)}, \quad (14)$$

$$X_0(19) : y^2 + y = x^3 + x^2 - 9x - 15 \quad \text{(Cremona label 19a1)}. \quad (15)$$

**Lemma 10.3.** Let $p = 5$. Then $\bar{\rho}_{E,\ell}$ is irreducible. Moreover, $\bar{\rho}_{E',\ell}$ is irreducible.

**Proof.** Suppose $\bar{\rho}_{E,\ell}$ is reducible. By the above there is an elliptic curve $E_2$ over the quadratic field $K = \mathbb{Q}(\sqrt{5})$, with a $K$-rational subgroup isomorphic to $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\ell\mathbb{Z}$. From classification of torsion subgroups of elliptic curves over quadratic fields [Kamienny 1992] we deduce that $\ell \leq 5$. However we are assuming throughout that $\ell \geq 5$ and $\ell \neq p$. This gives a contradiction as $p = 5$. Thus $\bar{\rho}_{E,\ell}$ is irreducible. The irreducibility of $\bar{\rho}_{E',\ell}$ follows from Lemma 10.1. \hfill \Box

**Lemma 10.4.** Let $p = 7$. Then $\bar{\rho}_{E,\ell}$ is irreducible.

---

$^1$At the suggestion of one of the referees we prove this statement. Let $P_1, P_2 \in E_1(K)$ be independent points of order 2. Let $Q$ be a solution to the equation $2X = P_1$. Then $Q$ has order 4 and the complete set of solutions is $\{Q, Q + P_2, 3Q, 3Q + P_2\}$, which is Galois-stable. Let $E_2 = E_1/\langle P_2 \rangle$ and let $\phi : E_1 \to E_2$ be the induced isogeny. As $\text{Ker}(\phi) \cap \langle Q \rangle = 0$, we see that $Q' = Q + \langle P_2 \rangle$ has order 4. Moreover, the set $\{Q', 3Q'\}$ is Galois-stable, so $\langle Q' \rangle$ is a $K$-rational cyclic subgroup of order 4 on $E_2$. The point of order $\ell$ on $E_1$ survives the isogeny, and so $E_2$ has a $K$-rational cyclic subgroup of order $4\ell$. 

We merely check that conditions (i)–(vi) of [Bruin and Najman 2016, Theorem 1] we need to deal with when applying that theorem. The theorem involves making certain choices and we indicate them briefly; in the notation of the theorem, we take $A = \mathbb{Z}/26\mathbb{Z}$, $L = \mathbb{Q}$, $m = 1$, $n = 26$, $X = X' = X_1(26)$, $p = p_0 = 7$. To apply the theorem we need the fact that the gonality of $X_1(26)$ is 6 [Derickx and van Hoeij 2014], and that its Jacobian has Mordell–Weil rank 0 over $\mathbb{Q}$ [Bruin and Najman 2016, page 11]. We merely check that conditions (i)–(vi) of [Bruin and Najman 2016, Theorem 1] are satisfied, and conclude that there are no elliptic curves over $K$ with a subgroup isomorphic to $\mathbb{Z}/26\mathbb{Z}$. This completes the proof.

**Lemma 10.5.** Let $p = 11$. Then $\tilde{\rho}_{E,\ell}$ is irreducible.

**Proof.** Now $K$ has degree 5. By the classification of cyclic $\ell$-torsion on elliptic curves over quintic fields [Derickx et al. ≥ 2016] we know that $\ell \leq 19$. As $\ell \neq p$ we need to deal with $\ell = 5, 7, 13, 17, 19$.

The elliptic curves $X_0(20), X_0(14)$ and $X_0(19)$ have rank 0 over $K$ and this allows us to quickly eliminate $\ell = 5, 7, 19$.

Suppose $\ell = 13$. We again apply [Bruin and Najman 2016, Theorem 1], with choices $A = \mathbb{Z}/26\mathbb{Z}$, $L = \mathbb{Q}$, $m = 1$, $n = 26$, $X = X' = X_1(26)$, $p = p_0 = 11$ (with Mordell–Weil and gonality information as in the proof of Lemma 10.4). This shows that there are no elliptic curves over $K$ with a subgroup isomorphic to $\mathbb{Z}/26\mathbb{Z}$, allowing us to eliminate $\ell = 13$.

Suppose $\ell = 17$. We apply the same theorem with choices $A = \mathbb{Z}/34\mathbb{Z}$, $L = \mathbb{Q}$, $m = 1$, $n = 34$, $X = X' = X_1(34)$, $p = p_0 = 11$. For this we need the fact that $X$ has gonality 10 [Derickx and van Hoeij 2014] and that the rank of $J_1(34)$ over $\mathbb{Q}$ is 0 [Bruin and Najman 2016, page 11]. Applying the theorem shows that there are no elliptic curves over $K$ with a subgroup isomorphic to $\mathbb{Z}/34\mathbb{Z}$. This completes the proof.

It remains to deal with $p = 13$. Unfortunately the field $K$ in this case is sextic, and the known bound [Derickx et al. ≥ 2016] for cyclic $\ell$-torsion over sextic fields is $\ell \leq 37$, and we have been unable to deal with the cases $\ell = 37$ directly over the sextic field. We therefore proceed a little differently. We are in fact most interested
in showing the irreducibility of $\tilde{\rho}_{E',\ell}$, where $E'$ is the Frey curve defined over the degree 3 subfield $K'$.

**Lemma 10.6.** Let $p = 13$. Then $\tilde{\rho}_{E',\ell}$ is irreducible.

**Proof.** Suppose $\tilde{\rho}_{E',\ell}$ is reducible. We treat the cases $13 \mid x$ and $13 \nmid x$ separately. Suppose first that $13 \mid x$. Then the curve $E'$ over the field $K'$ is semistable (Lemma 7.2). It is now straightforward to adapt the proof of Lemma 10.2 to show that $E'$ has nontrivial $\ell$-torsion, or is $\ell$-isogenous to an elliptic curve with nontrivial $\ell$-torsion. Thus there is an elliptic curve over $K'$ with a point of exact order $2\ell$. Now $K'$ is cubic, so by [Parent 2000; 2003] we have $\ell \leq 13$. As $\ell \neq p$, it remains to deal with the cases $\ell = 5, 7, 11$. The elliptic curves $X_0(14)$ and $X_0(11)$ have rank zero over $K'$, and in fact their $K'$-points are the same as their $\mathbb{Q}$-points. This easily allows us to eliminate $\ell = 7$ and $\ell = 11$ as before. The curve $X_0(10)$ has genus 0 so we need a different approach, and we leave this case to the end of the proof (recall that $E'$ does not necessarily have full 2-torsion over $K'$).

Now suppose that $13 \nmid x$. Here $E'/K'$ is not semistable. As we have assumed that $\tilde{\rho}_{E',\ell}$ is reducible, we have that $\tilde{\rho}_{E,\ell}$ is reducible (Lemma 10.1). Now we may apply Lemma 10.2 to deduce the existence of $E_1/K$ (which is $E$ or $\ell$-isogenous to it) that has a subgroup isomorphic to $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\ell\mathbb{Z}$. As before, let $p$ be the unique prime of $K$ above 13. By Lemma 6.1 the Frey curve $E$ has good reduction at $p$. As $p \nmid 2\ell$, we know from injectivity of torsion that $4\ell \mid \#E(F_p)$. But $F_p = F_{13}$. By the Hasse–Weil bounds,

$$\ell \leq (\sqrt{13} + 1)^2/4 \approx 5.3.$$ 

Thus $\ell = 5$.

It remains to deal with the case $\ell = 5$ for both $13 \mid x$ and $13 \nmid x$. In both cases we obtain a $K$-point on $X = X_0(20)$ whose image in $X_0(10)$ is a $K'$-point. We would like to compute $X(K)$. This computation proved beyond Magma’s capability. However, $K = K'(\sqrt{13})$. Thus the rank of $X(K)$ is the sum of the ranks of $X(K')$ and of $X'(K')$, where $X'$ is the quadratic twist of $X$ by 13. Computing the ranks of $X(K')$ and $X'(K')$ turns out to be a task within the capabilities of Magma, and we find that they are respectively 0 and 1. Thus $X(K)$ has rank 1. With a little more work we find that

$$X(K) = \frac{\mathbb{Z}}{6\mathbb{Z}} \cdot (4, 10) + \mathbb{Z} \cdot (3, 2\sqrt{13}).$$

Thus $X(K) = X(\mathbb{Q}(\sqrt{13}))$. It follows that the $j$-invariant of $E'$ must belong to $\mathbb{Q}(\sqrt{13})$. But the $j$-invariant belongs to $K'$ too, and so belongs to $\mathbb{Q}(\sqrt{13}) \cap K' = \mathbb{Q}$.

Let the rational integers $a, b$ be as in Sections 4 and 6. Recall that $b$ is odd, and that $\nu_2(a) = 5n$, where $n > 0$. Write $a = 2^{5n}a'$, where $a'$ is odd. We know that $\nu_2(j(E)) = -(20n - 4)$. The prime 2 is inert in $K'$. An explicit calculation,
making use of the fact that \( a' \equiv b \equiv 1 \pmod{2} \), shows that
\[
2^{20n-4} j(E) \equiv \frac{\theta_j^2 \theta_k^2}{(\theta_j - \theta_k)^2} \pmod{2}.
\]
Computing this residue for the possible values of \( j \) and \( k \), we check that it does not belong to \( \mathbb{F}_2 \), giving us a contradiction. \( \square \)

11. Proof of Theorem 1.1

In Section 5 we proved Theorem 1.1 for \( p = 3 \). In this section we deal with the values \( p = 5, 7, 11, 13 \). Let \( \ell, m \geq 5 \) be primes. Suppose \((x, y, z)\) is a primitive nontrivial solution to (2). Without loss of generality, \( 2 \mid x \). We let \( K = \mathbb{Q}(\zeta + \zeta^{-1}) \) where \( \zeta = \exp(2\pi i/p) \). For \( p = 13 \) we also let \( K' \) be the unique subfield of \( K \) of degree 3. Let \( E \) be the Frey curve attached to this solution \((x, y, z)\) defined in Section 6, where we take \( j = 1 \) and \( k = 2 \). For \( p = 13 \) we also work with the Frey curve \( E' \) defined in Section 7, where we take \( j = 1 \) and \( k = 5 \) (these choices satisfy the condition \( \tau(\theta_j) = \theta_k \), where \( \tau \) is unique involution on \( K \)). By Lemma 9.1 these elliptic curves are modular. Moreover, by the results of Section 10 the representation \( \bar{\rho}_{E,\ell} \) is irreducible for \( p = 5, 7, 11, 13 \), and the representation \( \bar{\rho}_{E',\ell} \) is irreducible for \( p = 13 \). Let \( \mathcal{K} \) be the number field \( K \) unless \( p = 13 \) and \( 13 \mid x \), in which case we take \( \mathcal{K} = K' \). Also let \( \mathcal{E} \) be the Frey curve \( E \) unless \( p = 13 \) and \( 13 \mid x \), in which we take \( \mathcal{E} \) to be \( E' \). By Lemma 8.3 there is a Hilbert cuspidal eigenform \( \mathfrak{f} \) over \( \mathcal{K} \) of parallel weight 2 and level \( \mathcal{N} \) as given in Table 1, such that \( \bar{\rho}_{E,\ell} \sim \bar{\rho}_{\mathfrak{f},\lambda} \), where \( \lambda \mid \ell \) is a prime of \( \mathbb{Q}_\mathfrak{f} \), the field generated by the Hecke eigenvalues of \( \mathfrak{f} \).

Using the Magma “Hilbert modular forms” package we compute the possible Hilbert newforms at these levels. The information is summarized in Table 1.

As shown in the table, there are no newforms at the relevant levels for \( p = 5 \), completing the contradiction for this case.\(^2\)

We now explain how we complete the contradiction for the remaining cases. Suppose \( q \) a prime of \( \mathcal{K} \) such that \( q \nmid 2 \ell \). In particular, \( q \) does not divide the level of \( \mathfrak{f} \), and \( \mathcal{E} \) has good or multiplicative reduction at \( q \). Write \( \sigma_q \) for a Frobenius element of \( G_{\mathcal{K}} \) at \( q \). Comparing the traces of \( \bar{\rho}_{\mathcal{E},\ell}(\sigma_q) \) and \( \bar{\rho}_{\mathfrak{f},\lambda}(\sigma_q) \) we obtain

(i) if \( \mathcal{E} \) has good reduction at \( q \) then \( a_q(\mathcal{E}) \equiv a_q(\mathfrak{f}) \pmod{\lambda} \);
(ii) if \( \mathcal{E} \) has split multiplicative reduction at \( q \) then \( \text{Norm}(q) + 1 \equiv a_q(\mathfrak{f}) \pmod{\lambda} \);
(iii) if \( \mathcal{E} \) has nonsplit multiplicative reduction at \( q \) then \(-\text{Norm}(q) + 1 \equiv a_q(\mathfrak{f}) \pmod{\lambda} \).

\(^2\)We point out in passing that for \( p = 5 \) we could have also worked with the Frey curve \( E'/\mathbb{Q} \). In that case the Hilbert newforms \( \mathfrak{f} \) are actually classical newforms of weight 2 and levels 2 and 50. There are no such newforms of level 2, but there are two newforms of level 50 corresponding to the elliptic curve isogeny classes 50a and 50b. These would require further work to eliminate.
Let $q \nmid 2p\ell$ be a rational prime and let

$$A_q = \{(\eta, \mu) : 0 \leq \eta, \mu \leq q - 1, \ (\eta, \mu) \neq (0, 0)\}.$$ 

For $(\eta, \mu) \in A_q$ let

$$u(\eta, \mu) = \begin{cases} 
(\theta_j + 2)\eta^2 + (\theta_j - 2)\mu^2 & \text{if } p \nmid x, \\
\frac{1}{(\theta_j - 2)}((\theta_j + 2)\eta^2 + (\theta_j - 2)\mu^2) & \text{if } p | x,
\end{cases}$$

and

$$v(\eta, \mu) = \begin{cases} 
-(\theta_k - 2)((\theta_k + 2)\eta^2 + (\theta_k - 2)\mu^2) & \text{if } p \nmid x, \\
\frac{1}{(\theta_k - 2)}((\theta_k + 2)\eta^2 + (\theta_k - 2)\mu^2) & \text{if } p | x.
\end{cases}$$

Write

$$E_{(\eta, \mu)} : Y^2 = X(X - u(\eta, \mu))(X + v(\eta, \mu)).$$

Let $\Delta(\eta, \mu)$, $c_4(\eta, \mu)$ and $c_6(\eta, \mu)$ be the usual invariants of this model. Let $\gamma(\eta, \mu) = -c_4(\eta, \mu)/c_6(\eta, \mu)$. Let $(a, b)$ be as in Section 4. As gcd$(a, b) = 1$, we have $(a, b) \equiv (\eta, \mu) \pmod{q}$ for some $(\eta, \mu) \in A_q$. In particular, $(a, b) \equiv (\eta, \mu) \pmod{q}$ for all primes $q | q$ of $K$. From the definitions of the Frey curves $E$ and $E'$ in Sections 6 and 7 we see that $E$ has good reduction at $q$ if and only if

<table>
<thead>
<tr>
<th>$p$</th>
<th>case</th>
<th>field $K$</th>
<th>Frey curve $E$</th>
<th>level $N$</th>
<th>eigenforms $f$</th>
<th>$[\mathbb{Q}_f : \mathbb{Q}]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>$5 \nmid x$</td>
<td>$K$</td>
<td>$E$</td>
<td>$2\mathcal{O}_K$</td>
<td>$-$</td>
<td>$-$</td>
</tr>
<tr>
<td></td>
<td>$5</td>
<td>x$</td>
<td>$K$</td>
<td>$E$</td>
<td>$2p$</td>
<td>$-$</td>
</tr>
<tr>
<td>7</td>
<td>$7 \nmid x$</td>
<td>$K$</td>
<td>$E$</td>
<td>$2\mathcal{O}_K$</td>
<td>$-$</td>
<td>$-$</td>
</tr>
<tr>
<td></td>
<td>$7</td>
<td>x$</td>
<td>$K$</td>
<td>$E$</td>
<td>$2p$</td>
<td>$f_1$</td>
</tr>
<tr>
<td>11</td>
<td>$11 \nmid x$</td>
<td>$K$</td>
<td>$E$</td>
<td>$2\mathcal{O}_K$</td>
<td>$f_2$</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>$11</td>
<td>x$</td>
<td>$K$</td>
<td>$E$</td>
<td>$2p$</td>
<td>$f_3, f_4$</td>
</tr>
<tr>
<td>13</td>
<td>$13 \nmid x$</td>
<td>$K$</td>
<td>$E$</td>
<td>$2\mathcal{O}_K$</td>
<td>$f_5, f_6$</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>$13</td>
<td>x$</td>
<td>$K$</td>
<td>$E$</td>
<td>$2\mathcal{O}_K$</td>
<td>$f_7, f_8$</td>
</tr>
<tr>
<td></td>
<td>$13 \mid x$</td>
<td>$K'$</td>
<td>$E'$</td>
<td>$2\mathfrak{B}$</td>
<td>$f_9, f_{10}$</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>$13</td>
<td>x$</td>
<td>$K'$</td>
<td>$E'$</td>
<td>$2\mathfrak{B}$</td>
<td>$f_{11}, f_{12}$</td>
</tr>
</tbody>
</table>

Table 1. Frey curve and Hilbert eigenform information. Here $p$ is the unique prime of $K$ above $p$, and $\mathfrak{B}$ is the unique prime of $K'$ above $p$. 


Weierstrass model which has conductor 26b1 with Cremona label \( \lambda \).

Table 2 gives our choices for the set \( S \). Then \( \lambda \mid B_q(f, \eta, \mu) \) gives a contradiction unless \( \lambda \mid \ell \), and we now show that it is indeed the case for \( \ell = 7 \). This completes the proof of Theorem 1.1.

The reader may be wondering whether we can eliminate the case \( p = 13 \) and \( \ell = 7 \) by enlarging our set \( S \); here we need only concern ourselves with forms \( f_9 \) and \( f_{11} \). Consider \( \eta, \mu = (0, 1) \), which belongs to \( A_q \) for any \( q \). The corresponding Weierstrass model \( E_{(0,1)} \) is singular with a split note. It follows that

\[
B_q(f, 0, 1) = \text{Norm}(q) + 1 - a_q(f).
\]

Note that if \( \lambda \) is a prime of \( \mathbb{Q}_f \) that divides \( \text{Norm}(q) + 1 - a_q(f) \) for all \( q \mid 26 \), then \( \ell \) will divide \( B_q(f) \) for any set \( S \) where \( \lambda \mid \ell \). This appears to be the case with \( \ell = 7 \) for \( f_{11} \), and we now show that it is indeed the case for \( f_9 \). Let \( F \) be the elliptic curve with Cremona label 26b1:

\[
F : y^2 + xy + y = x^3 - x^2 - 3x + 3,
\]

which has conductor 26b as an elliptic curve over \( K \). As \( K/\mathbb{Q} \) is cyclic, we know that \( F \) is modular over \( K \) and hence corresponds to a Hilbert modular form of parallel weight 2 and level 26b, and by comparing eigenvalues we can show that it in fact corresponds to \( f_9 \). Now the point \( (1, 0) \) on \( F \) has order 7. It follows that

\[
7 \mid \#E(\mathbb{F}_q) = \text{Norm}(q) + 1 - a_q(f) \quad \text{for all } q \mid 26,
\]

showing that for \( f_9 \) we can never
Table 2. Our choice of set of primes $S$ and the value of $B_S(f)$ for each of the eigenforms in Table 1.

<table>
<thead>
<tr>
<th>$p$</th>
<th>case</th>
<th>$S$</th>
<th>eigenform $f$</th>
<th>$B_S(f)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>7</td>
<td>{3}</td>
<td>$f_1$</td>
<td>$2^8 \times 3^5 \times 7^6$</td>
</tr>
<tr>
<td>11</td>
<td>11</td>
<td>{23, 43}</td>
<td>$f_2$</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>{23, 43}</td>
<td>$f_3$</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>{23, 43}</td>
<td>$f_4$</td>
<td>1</td>
</tr>
<tr>
<td>13</td>
<td>13</td>
<td>{79, 103}</td>
<td>$f_5$</td>
<td>$2^{6240} \times 3^{12}$</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>{79, 103}</td>
<td>$f_6$</td>
<td>$2^{12792} \times 3^{24}$</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>{79, 103}</td>
<td>$f_7$</td>
<td>$2^{10608} \times 3^{64}$</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>{79, 103}</td>
<td>$f_8$</td>
<td>$2^{18720} \times 3^{936}$</td>
</tr>
<tr>
<td>13</td>
<td>13</td>
<td>{3, 5, 31, 47}</td>
<td>$f_9$</td>
<td>$7^2$</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>{3, 5, 31, 47}</td>
<td>$f_{10}$</td>
<td>$3^7$</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>{3, 5, 31, 47}</td>
<td>$f_{11}$</td>
<td>$7^6$</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>{3, 5, 31, 47}</td>
<td>$f_{12}$</td>
<td>1</td>
</tr>
</tbody>
</table>

We can eliminate $\ell = 7$ by enlarging the set $S$. We can still complete the contradiction in this case as follows. Note that $\tilde{\rho}_{f_9, 7} \sim \tilde{\rho}_{F, 7}$ which is reducible. As $\tilde{\rho}_{E, 7}$ is irreducible we have $\tilde{\rho}_{E, 7} \not\sim \tilde{\rho}_{f_9, 7}$, completing the contradiction for $f = f_9$. We strongly suspect that reducibility of $\tilde{\rho}_{f_{11}, \lambda}$ (where $\lambda$ is the unique prime above 7 of $\mathbb{Q}_{f_{11}}$), but we are unable to prove it.

Remark. We now explain why we believe that the above strategy will succeed in proving that (2) has no nontrivial primitive solutions, or at least in bounding the exponent $\ell$, for larger values of $p$ provided the eigenforms $f$ at the relevant levels can be computed. The usual obstruction to bounding the exponent (see [Siksek 2012, Section 9]) comes from eigenforms $f$ that correspond to elliptic curves with a torsion structure that matches the Frey curve $E$. Let $f$ be such an eigenform. Let $q \nmid 2p$ be a rational prime and $q_1, \ldots, q_r$ be the primes of $\mathcal{K}$ above $q$. Note that $\text{Norm}(q_1) = \cdots = \text{Norm}(q_r) = q^{d/r}$, where $d = [\mathcal{K} : \mathbb{Q}]$. We would like to estimate the “probability” that $B_q(f)$ is nonzero. Observe that if $B_q(f)$ is nonzero, then we obtain a bound for $\ell$. Examining the definitions above shows that the ideal $B_q(f)$ is 0 if and only if there is some $(\eta, \mu) \in A_q$ such that $a_q(E(\eta, \mu)) = a_q(f)$ for $q = q_1, q_2, \ldots, q_r$. Treating $a_q(E(\eta, \mu))$ as a random variable belonging to the Hasse interval $[-2q^{d/(2r)}, 2q^{d/(2r)}]$, we see that the “probability” that $a_q(E(\eta, \mu)) = a_q(f)$ is roughly $c/q^{d/(2r)}$, with $c = 1/4$.

We can be a little more sophisticated and take account of the fact that the torsion structures coincide, and that these impose congruence restrictions on both
traces. In that case we should take $c = 1$ if $E$ has full 2-torsion (i.e., $E$ is the Frey curve $E$) and take $c = \frac{1}{2}$ if $E$ has just one nontrivial point of order 2 (i.e., $E = E'$ and $p \equiv 1 \pmod{4}$). Thus the “probability” that $a_q(E_{(\eta, \mu)}) = a_q(\mathfrak{f})$ for all $q \mid q$ simultaneously is roughly $c^r / q^{d/2}$. Since $B_q(\mathfrak{f}) = q \prod_{(\eta, \mu) \in A_q} B_q(\mathfrak{f}, \eta, \mu)$, it follows that the “probability” $P_q$ (say) that $B_q(\mathfrak{f})$ is nonzero satisfies

$$P_q \sim \left(1 - \frac{c^r}{q^{d/2}}\right)^{q^{d/2} - 1}.$$

For $q$ large, we have $(1 - c^r / q^{d/2})^{q^{d/2}} \approx e^{-c^r}$. For $d \geq 5$, from the above estimates, we expect that $P_q \to 1$ as $q \to \infty$. Thus we certainly expect our strategy to succeed in bounding the exponent $\ell$.

Acknowledgments

We are grateful to the three referees for their careful reading of the paper and for suggesting many improvements. We are indebted to Lassina Dembélé, Steve Donnelly, Marc Masdeu and Jack Thorne for stimulating conversations.

References


Communicated by Joseph H. Silverman
Received 2015-06-09 Revised 2016-03-22 Accepted 2016-06-22

samuele.anni@gmail.com Mathematics Institute, University of Warwick, Coventry CV4 7AL, United Kingdom

samir.siksek@gmail.com Mathematics Institute, University of Warwick, Coventry CV4 7AL, United Kingdom
We explore the geometry and establish the slope-stability of tautological vector bundles on Hilbert schemes of points on smooth surfaces. By establishing stability in general, we complete a series of results of Schlickewei and Wandel, who proved the slope-stability of these vector bundles for Hilbert schemes of 2 points or 3 points on K3 or abelian surfaces with Picard group restrictions. In exploring the geometry, we show that every sufficiently positive semistable vector bundle on a smooth curve arises as the restriction of a tautological vector bundle on the Hilbert scheme of points on the projective plane. Moreover, we show that the tautological bundle of the tangent bundle is naturally isomorphic to the log tangent sheaf of the exceptional divisor of the Hilbert–Chow morphism.

Introduction

The purpose of this paper is to explore the geometry of tautological bundles on Hilbert schemes of smooth surfaces and to establish the slope-stability of these bundles.

Let $S$ be a smooth complex projective surface, and denote by $S^{[n]}$ the Hilbert scheme parametrizing length-$n$ subschemes of $S$. This parameter space carries some natural tautological vector bundles: if $L$ is a line bundle on $S$ then $L^{[n]}$ is the rank-$n$ vector bundle whose fiber at the point corresponding to a length-$n$ subscheme $\xi \subset S$ is the vector space $H^0(S, L \otimes O_\xi)$. These tautological vector bundles have attracted a great deal of interest. Lehn [1999] first computed the cohomology of the tautological bundles. Later Danila [2001] and Scala [2009] identified the induced symmetric group representations on the cohomology of the tautological bundles. Ellingsrud and Strømme [1993] showed that the Chern classes of the bundles $O_{P^2}^{[n]}$, $O_{P^2}(1)^{[n]}$, and $O_{P^2}(2)^{[n]}$ generate the cohomology of $(P^2)^{[n]}$. Nakajima gave a nicely exposited interpretation [1999, §4.3] of the McKay correspondence by

MSC2010: 14J60.

Keywords: Hilbert schemes of surfaces, vector bundles on surfaces, Fourier–Mukai transforms, slope-stability, spectral curves, log tangent bundle, tautological bundles, Hilbert schemes of points.
restricting the tautological bundles to the $G$-Hilbert scheme. Recently Okounkov [2014] formulated a conjecture about special generating functions associated to the tautological bundles.

Given the importance of the tautological bundles, it is natural to explore how different geometric aspects of vector bundles transform to their tautological bundles. For instance, we ask when the tautological bundle of a stable bundle is also stable. In [Schlickewei 2010; Wandel 2013; 2014] this question has been answered positively for Hilbert schemes of 2 points or 3 points on a K3 or abelian surface with Picard group restrictions. Our first result establishes the stability of these bundles for arbitrary $n$ and any surface.

**Theorem A.** If $\mathcal{L}$ is a nontrivial line bundle on $S$, then $\mathcal{L}^{[n]}$ is slope-stable with respect to natural Chow divisors on $S^{[n]}$.

More precisely, an ample divisor on $S$ determines a natural ample divisor on $\text{Sym}^n(S)$, and the pullback via the Hilbert–Chow morphism gives one such natural Chow divisor on $S^{[n]}$, which is not ample but is big and semiample. More generally, we prove that if $\mathcal{E} \not\cong \mathcal{O}_S$ is any slope-stable vector bundle on $S$ with respect to some ample divisor then $\mathcal{E}^{[n]}$ is slope-stable with respect to the corresponding Chow divisor. Although Theorem A only gives stability with respect to a strictly big and nef divisor, we are able to deduce stability with respect to nearby ample divisors via a perturbation argument on the nef cone.

If $S$ is any smooth surface, there is a divisor $B_n$ in $S^{[n]}$ which consists of nonreduced subschemes. The pair $(S^{[n]}, B_n)$ gives a natural closure of the space of $n$ distinct points in $S$. The vector fields on $S^{[n]}$ tangent to $B_n$ form the sheaf of logarithmic vector fields $\text{Der}_C(-\log B_n)$. Our second result says the sheaf $\text{Der}_C(-\log B_n)$ is naturally isomorphic to the tautological bundle associated to the tangent bundle on $S$.

**Theorem B.** For any smooth surface $S$ there exists a natural injection

$$\alpha_n : (T_S)^{[n]} \rightarrow T_{S^{[n]}},$$

and $\alpha_n$ induces an isomorphism between $(T_S)^{[n]}$ and $\text{Der}_C(-\log B_n)$.

The analogous statement also holds for smooth curves. In general, the sheaves $\text{Der}_C(-\log B_n)$ are only guaranteed to be reflexive, as $B_n$ is not a simple normal crossing divisor. However, Theorem B shows $\text{Der}_C(-\log B_n)$ is locally free; that is, $B_n$ is a free divisor. Buchweitz, Ebeling, and Graf von Bothmer [Buchweitz et al. 2009] have already shown that $B_n$ is a free divisor using different methods.

Using Aubin and Yau’s theorem [Aubin 1976] we obtain:

**Corollary C.** If a surface $S$ has ample canonical bundle, then the log tangent bundle $\text{Der}_C(-\log B_n)$ is polystable with respect to the big and nef canonical divisor $K_{S^{[n]}}$.
Finally, we explore the geometry of the tautological bundles when the surface is the projective plane. We prove that the tautological bundles on $(\mathbb{P}^2)^{[n]}$ are rich enough to capture all semistable rank-$n$ bundles on curves.

**Theorem D.** If $C$ is a smooth projective curve and $E$ is a semistable rank-$n$ vector bundle on $C$ with sufficiently positive degree, then there exists an embedding $C \to (\mathbb{P}^2)^{[n]}$ such that

$$O_{\mathbb{P}^2}(1)^{[n]}|_C \cong E.$$ 

The proof of Theorem A follows the approach taken by Mistretta [2006], who studies the stability of tautological bundles on the symmetric powers of a curve. The idea is to examine the tautological vector bundles on the cartesian power $S^n$ and show there are no $S_n$-equivariant destabilizing subsheaves. This strategy is more effective for surfaces because the diagonals in $S^n$ have codimension 2. The map in Theorem B arises from pushing forward the normal sequence of the universal family. The proof of Theorem D is constructive, using the spectral curves of Beauville, Narasimhan, and Ramanan [Beauville et al. 1989].

In Section 1 we give the proof of Theorem A. In Section 2 we prove Theorem B and deduce Corollary C. In Section 3 we prove Theorem D. In Section 4 we give the perturbation argument, deducing that the tautological bundles are stable with respect to ample divisors.

Throughout, we work over the complex numbers. If $X$ is a variety of dimension $d$ and $E$ is a vector bundle on $X$, then for any divisor class $H \in N^1(X)$ we define the slope of $E$ with respect to $H$ to be the rational number

$$\mu_H(E) := \frac{c_1(E) \cdot H^{d-1}}{\text{rank}(E)}.$$ 

We say $E$ is slope-stable (resp. slope-semistable) with respect to $H$ if, for all subsheaves $F \subset E$ of intermediate rank, we have

$$\mu_H(F) < \mu_H(E) \quad (\text{resp.} \ \mu_H(F) \leq \mu_H(E)).$$ 

1. Stability of tautological bundles

In this section we prove that the tautological bundle of a stable vector bundle $E$ is stable with respect to natural Chow divisors on $S^{[n]}$. Thus we deduce Theorem A when $E$ is a nontrivial line bundle. We start by defining the essential objects in the study of Hilbert schemes of points on surfaces.

Let $S$ be a smooth complex projective surface. We write $S^{[n]}$ for the Hilbert scheme of length-$n$ subschemes of $S$. We denote by $\mathcal{Z}_n$ the universal family of $S^{[n]}$.
with the following projections:

\[
\begin{array}{ccc}
S \times S^{[n]} & \supset & Z_n \\
p_1 & \mapsto & S \\
p_2 & \downarrow & S^{[n]}
\end{array}
\]

For a fixed vector bundle \( E \) on \( S \) of rank \( r \), we define

\[
E^{[n]} := (p_2)_*(p_1^*E),
\]

which is the tautological vector bundle associated to \( E \) and has rank \( rn \). The fiber of \( E^{[n]} \) at a point \([\xi] \in S^{[n]}\) can be naturally identified with the vector space \( H^0(S, E|_{[\xi]})\).

The symmetric group on \( n \) elements, \( S_n \), naturally acts on the cartesian product \( S^n \), and we write \( \sigma_n \) for the quotient map

\[
\sigma_n : S^n \to S^n / S_n =: \text{Sym}^n(S).
\]

There is also a Hilbert–Chow morphism,

\[
h_n : S^{[n]} \to \text{Sym}^n(S),
\]

which is a semismall map [de Cataldo and Migliorini 2002, Definition 2.1.1].

We wish to view \( E^{[n]} \) as an \( S_n \)-equivariant sheaf on \( S^n \). Recall that if \( G \) is a finite group that acts on a scheme \( X \), and if \( F \) is a coherent sheaf on \( X \), then a \( G \)-equivariant structure on \( F \) is given by a choice of isomorphisms \( \phi_g : F \to g^*F \) for all \( g \in G \) satisfying the compatibility condition \( h^*(\phi_g) \circ \phi_n = \phi_{gh} \). Following [Danila 2001] and [Scala 2009], we study the tautological bundles on \( S^{[n]} \) by working with \( S_n \)-equivariant sheaves on \( S^n \). For our purposes it is enough to study \( E^{[n]} \) equivariantly on the open subset of distinct points in \( S^{[n]} \).

We write \( \text{Sym}^n(S)_\circ \) for the open subset of \( \text{Sym}^n(S) \) of distinct points. Likewise, given a map \( f : X \to \text{Sym}^n(S) \), we write \( X_\circ \) for \( f^{-1}(\text{Sym}^n(S)_\circ) \). By abuse of notation, given another map \( g : X \to Y \) with domain \( X \) we define \( g_\circ := g|_{X_\circ} \), and given a coherent sheaf \( F \) on \( X \) we define \( F_\circ := F|_{X_\circ} \). The map \( h_{n,\circ} : S^{[n]}_\circ \to \text{Sym}^n(S)_\circ \) is an isomorphism. We define

\[
\bar{\sigma}_{n,\circ} := h_{n,\circ}^{-1} \circ \sigma_{n,\circ} : S^n_\circ \to S^{[n]}_\circ.
\]

Given a torsion-free coherent sheaf \( F \) on \( S^{[n]} \), we define a torsion-free coherent sheaf on \( S^n \) by

\[
(F)_{S^n} := j_*(\bar{\sigma}_{n,\circ}^*(F_\circ))
\]

where \( j \) is the inclusion \( j : S^n_\circ \to S^n \). The sheaf \( (F)_{S^n} \) can be thought of as a modification of \( F \) along the exceptional divisor of \( h_n \).
The pullback $\overline{\sigma}_{n,o}^* (-)$ is left exact, as the map $\overline{\sigma}_{n,o}$ is étale; thus the functor $(-)_{S^n}$ is left exact. If $F$ is reflexive, the normality of $S^n$ implies that the natural $\mathfrak{S}_n$-equivariant structure on the reflexive sheaf $\overline{\sigma}_{n,o}^*(F_o)$ pushes forward uniquely to an $\mathfrak{S}_n$-equivariant structure on $(F)_{S^n}$.

Let $q_i$ denote the projection from $S^n$ onto the $i$-th factor. Given a vector bundle $E$ on $S$, there is an $\mathfrak{S}_n$-equivariant vector bundle on $S^n$ defined by

$$E^\boxplus n := \bigoplus_{i=1}^n q_i^*(E).$$

We have given two natural $\mathfrak{S}_n$-equivariant sheaves on $S^n$ associated to $E$. In fact, they are equivalent.

**Lemma 1.1.** Given a vector bundle $E$ on $S$ there is an isomorphism

$$(E[n])_{S^n} \cong E^\boxplus n$$

of $\mathfrak{S}_n$-equivariant vector bundles on $S^n$.

**Proof.** Consider the following fiber square:

$$F := \mathcal{Z}_{n,o} \times_{S^n} S^n_o \xrightarrow{p^\prime_{2,o}} S^n_o \xrightarrow{\overline{\sigma}_{n,o}^\prime} \mathcal{Z}_{n,o} \xrightarrow{p_{2,o}} S^n_o \xrightarrow{\overline{\sigma}_{n,o}} S^n_{[n]}.$$ 

Every map in the fiber square is an étale map between $\mathfrak{S}_n$-schemes (the $\mathfrak{S}_n$-action on $\mathcal{Z}_{n,o}$ and $S^n_o$ is trivial). We write $\Gamma_i$ for the subscheme of $S^n_o \times S$ that is the graph of the map $q_{i,o} : S^n_o \to S$. The scheme $F$ is equal to the disjoint union $\bigsqcup \Gamma_i$ and is a subscheme of $S^n_o \times S$. The restriction $p^\prime_{1,o} \circ \overline{\sigma}_{n,o}^\prime |_{\Gamma_i}$ is the projection $\Gamma_i \to S$. So there is an equivariant isomorphism

$$(p^\prime_{2,o})_*((\overline{\sigma}_{n,o}^\prime)^*(p^\prime_{1,o}^*(E))) \cong \overline{\sigma}_{n,o}^*(E^\boxplus n).$$

As the fiber square is made of flat proper $\mathfrak{S}_n$-maps, there is a natural $\mathfrak{S}_n$-equivariant isomorphism

$$(p^\prime_{2,o})_*((\overline{\sigma}_{n,o}^\prime)^*(p^\prime_{1,o}^*(E))) \cong \overline{\sigma}_{n,o}^*( (p^\prime_{2,o})_*((\overline{\sigma}_{n,o}^\prime)^*(p^\prime_{1,o}^*(E)))).$$ 

The latter sheaf is $(E[n])_{S^n_o}$. Finally, any isomorphism between vector bundles on $S^n_o$ uniquely extends to an isomorphism between their pushforwards along $j$. Therefore, there is a natural $\mathfrak{S}_n$-equivariant isomorphism $(E[n])_{S^n} \cong E^\boxplus n$. □
Given an ample divisor \( H \) on \( S \), there is a natural \( S^n \)-invariant ample divisor on \( S^n \) defined by

\[
H_{S^n} := \sum_{i=1}^{n} q_i^*(H).
\]

This is the Chow divisor that appears in Theorem A. Fogarty [1973, Lemma 6.1] shows every divisor \( H_{S^n} \) descends to an ample Cartier divisor on \( \text{Sym}^n(S) \). Pulling back this Cartier divisor along the Hilbert–Chow morphism gives a big and nef divisor on \( S[n] \), which we denote by \( H_n \). If \( H \) is effective then \( H_n \) can be realized set-theoretically as

\[
H_n = \{ \xi \in S[n] \mid \xi \cap \text{Supp}(H) \neq \emptyset \}.
\]

**Lemma 1.2.** If \( \mathcal{F} \) is a torsion-free sheaf on \( S[n] \) then

\[
(n!) \int_{S[n]} c_1(\mathcal{F}) \cdot (H_n)^{2n-1} = \int_{S^n} c_1((\mathcal{F})_{S^n}) \cdot (H_{S^n})^{2n-1}.
\]

**Proof.** This is a straightforward calculation using \( S[n] , \text{Sym}^n(S)_o , \) and \( S^n_o \). \( \square \)

In the following lemma we assume Proposition 4.7, which says the pullback of a stable bundle to a product is stable with respect to a product polarization. For the sake of the exposition we give the proof of Proposition 4.7 in Section 4.

**Lemma 1.3.** If \( \mathcal{E} \not\cong \mathcal{O}_S \) is slope-stable on \( S \) with respect to an ample divisor \( H \) then there are no \( S^n \)-equivariant subsheaves of \( \mathcal{E}^{\oplus n} \) that are slope-destabilizing with respect to \( H_{S^n} \).

**Proof.** Let \( 0 \neq \mathcal{F} \subset \mathcal{E}^{\oplus n} \) be an \( S^n \)-equivariant subsheaf. We can find a (not necessarily equivariant) slope-stable subsheaf \( 0 \neq \mathcal{F}' \subset \mathcal{F} \) which has maximal slope with respect to \( H_{S^n} \). Fix \( i \) so that the composition

\[
\mathcal{F}' \to \mathcal{E}^{\oplus n} \to q_i^* \mathcal{E}
\]

is nonzero. By Proposition 4.7 we know that each \( q_i^* \mathcal{E} \) is slope-stable with respect to \( H_{S^n} \). A nonzero map between slope-stable sheaves can only exist if

1. the slope of \( \mathcal{F}' \) is less than the slope of \( q_i^* \mathcal{E} \), or
2. \( \mathcal{F}' \to q_i^* \mathcal{E} \) is an isomorphism.

In case (1), \( \mu_{H_{S^n}}(\mathcal{F}) \leq \mu_{H_{S^n}}(\mathcal{F}') < \mu_{H_{S^n}}(q_i^* \mathcal{E}) \). By symmetry, \( \mu_{H_{S^n}}(q_i^* \mathcal{E}) = \mu_{H_{S^n}}(q_j^* \mathcal{E}) \) for all \( i \) and \( j \). Thus we have \( \mu_{H_{S^n}}(q_i^* \mathcal{E}) = \mu_{H_{S^n}}(\mathcal{E}^{\oplus n}) \), and \( \mathcal{F} \) does not destabilize \( \mathcal{E}^{\oplus n} \).
In case (2), we know $F' \cong q_i^* E$. Because $E \not\cong O_S$, the pullbacks $q_i^* E$ and $q_j^* E$ are not isomorphic unless $i = j$. As all the $q_j^* E$ have the same slope and are stable with respect to $H_{Sn}$, we have $\text{Hom}(F', q_j^* E) = 0$ for $j \neq i$. In particular, all the compositions

$$F' \to E \cong q_j^* E$$

are zero for $j \neq i$. Thus $F'$ is a summand of $E \cong q_j^* E$. So $F$ is an $S_n$-equivariant subsheaf of $E \cong q_j^* E$, which contains one of the summands. But $S_n$ acts transitively on the summands so $F$ contains all the summands, hence $F$ does not destabilize $E \cong q_j^* E$. □

Now we prove Theorem A in full generality.

**Theorem 1.4.** If $E \not\cong O_S$ is a vector bundle on $S$ which is slope-stable with respect to an ample divisor $H$, then $E[n]$ is slope-stable with respect to $H_n$.

**Proof.** Let $F \subset E[n]$ be a reflexive subsheaf of intermediate rank. It is enough to consider reflexive sheaves because the saturation of a torsion-free subsheaf of $E[n]$ is reflexive of the same rank and its slope cannot decrease. By Lemma 1.2, the slope of a torsion-free sheaf $F$ with respect to $H_n$ is, up to a fixed positive multiple, the same as the slope of $(F)_{S^n}$ with respect to $H_{Sn}$. In particular,

$$\mu_{H_n}(F) < \mu_{H_n}(E[n]) \iff \mu_{H_{Sn}}((F)_{S^n}) < \mu_{H_{Sn}}(E \cong q_j^* E)$$.

Now $(F)_{S^n}$ is naturally an $S_n$-equivariant subsheaf of $E \cong q_j^* E$. Thus, by Lemma 1.3,

$$\mu_{H_{Sn}}((F)_{S^n}) < \mu_{H_{Sn}}(E \cong q_j^* E)$$.

Therefore, $\mu_{H_n}(F) < \mu_{H_n}(E[n])$ for all torsion-free subsheaves of intermediate rank, and $E[n]$ is stable with respect to $H_n$. □

2. **The tautological tangent map**

For any smooth (not necessarily projective) surface $S$, the Hilbert scheme $S[n]$ is a smooth closure of the space of $n$ distinct points in $S$. The boundary $B_n$ is the locus of nonreduced length-$n$ subschemes of $S$. We are interested in vector fields which are tangent to the boundary $B_n$.

**Definition 2.1.** If $D$ is a codimension-1 subvariety of a smooth variety $X$, then the sheaf of logarithmic vector fields, denoted $\text{Der}_C(-\log D)$, is the subsheaf of $T_X$ consisting of vector fields which along the regular locus of $D$ are tangent to $D$.

When $D$ is smooth, $\text{Der}_C(-\log D)$ is just the elementary transformation of the tangent bundle along the normal bundle of $D$ in $X$; in particular, it is a vector bundle. Even when $D$ is singular, $\text{Der}_C(-\log D)$ is reflexive by definition, so it is enough to define $\text{Der}_C(-\log D)$ away from the singular locus (or any codimension-2 set in $X$) of $D$ and then pushforward.
For Hilbert schemes of points on a surface, \( \text{Der}_C(-\log B_n) \) can be naturally understood as the tautological bundle of the tangent bundle on the surface.

**Theorem B.** For any smooth connected surface \( S \) there exists a natural injection

\[
\alpha_n : (T_S)^{[n]} \to T_{S^{[n]}},
\]

and \( \alpha_n \) induces an isomorphism between \( (T_S)^{[n]} \) and \( \text{Der}_C(-\log B_n) \).

At a point \([\xi] \in S^{[n]}\) the map \( \alpha_n|_{[\xi]} \) can be interpreted as deformations of \( \xi \) coming from tangent vectors of \( S \). We expect that the degeneracy loci of \( \alpha_n \) give an interesting stratification of \( S^{[n]} \).

Before proving Theorem B we prove a general lemma.

**Lemma 2.2.** Let \( X \) and \( Y \) be smooth varieties and \( f : X \to Y \) a branched covering with reduced branch locus \( B \subset Y \). If \( \delta \in H^0(Y, TY) \) is a vector field on \( Y \) whose pullback \( f^*\delta \in H^0(X, f^*TY) \) is in the image of

\[
d f : H^0(X, TX) \to H^0(X, f^*TY),
\]

then \( \delta \in H^0(Y, \text{Der}_C(-\log B)) \).

**Proof.** It is enough to check that \( \delta \) is tangent to \( B \) for points \( p \in B \) outside of a codimension-2 subset in \( Y \). Let \( p \in B \) be a general point and \( q \) a ramified point in the fiber of \( f \) over \( p \). We can choose local analytic coordinates \( y_1, \ldots, y_n \) centered at \( p \) and coordinates \( x_1, \ldots, x_n \) centered at \( q \) such that

\[
\begin{align*}
    f^*(y_1) &= x_1^m, \\
    f^*(y_i) &= x_i & \text{for } i > 1.
\end{align*}
\]

That is, \( y_1 \) is a local equation for \( B \) and \( x_1 \) is a local equation for the reduced component of ramification containing \( q \). Then the derivative \( df \) maps

\[
\frac{\partial}{\partial x_1} \mapsto m x_1^{m-1} f^* \left( \frac{\partial}{\partial y_1} \right), \quad \frac{\partial}{\partial x_i} \mapsto f^* \left( \frac{\partial}{\partial y_i} \right) \quad \text{for } i > 1.
\]

Now \( f^*\delta \) is in the image of \( df \). Expanding locally,

\[
f^*\delta = f^*(g_1) f^* \left( \frac{\partial}{\partial y_1} \right) + \cdots + f^*(g_n) f^* \left( \frac{\partial}{\partial y_n} \right).
\]

Thus \( x_1^{m-1} \) divides \( f^*(g_1) \). So \( y_1 \) divides \( g_1 \) and \( \delta \) is in \( H^0(Y, \text{Der}_C(-\log B)) \). \( \square \)

**Proof of Theorem B.** As in Section 1 we use \( Z_n \subset S \times S^{[n]} \) to denote the universal family of the Hilbert scheme of points. Applying relative Serre duality to the main result of [Lehn 1998] shows that the tangent bundle of \( S^{[n]} \) is given by \( T_{S^{[n]}} = (p_2)_* \mathcal{Hom}(I_{Z_n}, \mathcal{O}_{Z_n}) \). The normal sequence for \( Z_n \) gives a map

\[
p_1^* T_S \oplus p_2^* T_{S^{[n]}} \cong T_{S \times S^{[n]}}|_{Z_n} \xrightarrow{\beta} (I_{Z_n}/I_{Z_n}^2)^\vee \cong \mathcal{Hom}(I_{Z_n}, \mathcal{O}_{Z_n}).
\]
Thus after pushing forward the first summand we get a map

\[ \alpha_n : (T_S)[n] := (p_2)_*(p_1^*T_S) \rightarrow (p_2)_* \mathcal{H}om(\mathcal{I}_{Z_n}, \mathcal{O}_{Z_n}) = T_{S^n}. \]

To prove that \( \alpha_n \) maps \((T_S)[n]\) isomorphically onto \( \text{Der}_{\mathbb{C}}(-\log B_n) \) we first restrict to the open set \( U \subset S^{[n]} \) parametrizing subschemes \( \xi \subset S \), where \( \xi \) contains at least \( n - 1 \) distinct points. The complement of \( U \) has codimension 2 so by reflexivity it is enough to prove the theorem on \( U \). Moreover, the open set

\[ V := p_2^{-1}U \subset Z_n \]

is smooth so we are in a situation where we can apply Lemma 2.2. There is a map

\[ 0 \rightarrow T_{Z_n}|_V \rightarrow p_2^*(T_S[n])|_V \oplus p_1^*T_S|_V \xrightarrow{\beta} \mathcal{H}om(\mathcal{I}_{Z_n}, \mathcal{O}_{Z_n})|_V \]

in which \( \phi \) is the natural map coming from pulling back a pushforward. The composition

\[ \beta \circ (p_2^*\alpha_n|_V \oplus -\phi|_V) \]

is identically zero. Therefore, the pullback of each local section of \((T_S)[n]|_U\) lies in \( T_{Z_n}|_V \). It follows from Lemma 2.2 that \((T_S)[n]\) is contained in \( \text{Der}_{\mathbb{C}}(-\log B_n) \). Now we can think of \( \alpha_n \) as having codomain \( \text{Der}_{\mathbb{C}}(-\log B_n) \). The map is an isomorphism of \((T_S)[n]\) and \( \text{Der}_{\mathbb{C}}(-\log B_n) \) away from \( B_n \) and they both have the same first Chern class. Therefore, \( \alpha_n \) could only fail to be an isomorphism in codimension greater than 2. But both sheaves are reflexive, and any isomorphism between reflexive sheaves away from codimension 2 on a normal variety extends uniquely to an isomorphism on the whole variety. \( \square \)

**Proof of Corollary C.** As a reminder, a vector bundle is polystable if it is a direct sum of stable bundles of the same slope. The theorem of Aubin and Yau [Aubin 1976] proves the existence of Kähler–Einstein metrics for canonically polarized manifolds. This implies that the tangent bundle is polystable with respect to the canonical bundle (see [Kobayashi 1987, Theorem 8.3]; this is the easy direction of the Donaldson–Uhlenbeck–Yau theorem [Donaldson 1985]). Thus \( T_S \) is either stable or a direct sum of line bundles of the same canonical degree. In the first case, Corollary C follows directly from Theorems A and B.

For the second case, let \( T_S \cong \mathcal{L}_1 \oplus \mathcal{L}_2 \). First we point out that taking tautological bundles respects direct sums; that is,

\[ (\mathcal{E} \oplus \mathcal{F})^{[n]} \cong \mathcal{E}^{[n]} \oplus \mathcal{F}^{[n]}. \]
We then note that neither $L_1$ nor $L_2$ is trivial so their tautological bundles are stable by Theorem A. And if two line bundles on $S$ have equal degrees with respect to the canonical bundle then their tautological bundles also have equal degrees with respect to $K_{S[n]}$. Thus, by Theorem B, $\text{Der}_C(−\log B_n)$ is a direct sum of stable bundles of the same slope with respect to $K_{S[n]}$, proving Corollary C. □

Remark 2.3 (on the rank of $\alpha_n$). The restriction of $\alpha_n$ to any point $[ξ] ∈ S[n]$ is precisely the map from $H^0(S, T_S|_ξ)$ to $\text{Hom}(I_ξ, O_ξ)$ in the normal sequence of $ξ ⊂ S$. In [Bejleri and Stapleton 2016] we relate the rank of $\alpha_n$ to the dimension of the tangent space of the fibers of the Hilbert–Chow morphism. In particular, we show that if $ξ ⊂ C^2$ is cut out by monomials and $P_ξ$ denotes the fiber of the Hilbert–Chow morphism at $ξ$, then

$$\dim T_{[ξ]} P_ξ = 2n − \text{rank}(\alpha_n|_{[ξ]}).$$

Moreover, we give an explicit combinatorial formula for computing $\text{rank}(\alpha_n|_{[ξ]})$ at these monomial subschemes.

3. Spectral curves and tautological bundles

In this section we prove that every sufficiently positive, rank-$n$, semistable vector bundle on a smooth projective curve arises as the pullback of $O_{P^2}(1)|n$ along an embedding of the curve in $(P^2)^{|n|}$. To prove the theorem we need the spectral curves of [Beauville et al. 1989]. For completeness, we recall the construction.

Let $π : D → C$ be an $n:1$ map between smooth irreducible projective curves and let $E$ be an $O_C$-module. If $D$ can be embedded into the total space

$$\mathbb{L} := \text{Spec}_{O_C}(\text{Sym}^*(L^\vee)) \xrightarrow{π_\pi} C$$

of a line bundle $L$ on $C$, with $π = π_{\mathbb{L}}|_D$, then this gives a presentation

$$π_\pi O_D ≅ \text{Sym}^*(L^\vee)/(x^n + s_1 x^{n-1} + ⋯ + s_n)$$

for $x^n + s_1 x^{n-1} + ⋯ + s_n ∈ H^0(\mathbb{L}, (π_{\mathbb{L}}^* L)^\otimes n)$. Here we write $x ∈ H^0(\mathbb{L}, π_{\mathbb{L}}^* (L))$ for the coordinate section of $π_{\mathbb{L}}^* (L)$. To give $E$ the structure of a $π_\pi O_D$-module we need to specify a multiplication map $m : E ⊗ L^{-1} → E$ (equivalently $E → E ⊗ L$) which satisfies the relation $m^n + s_1 m^{n-1} + ⋯ + s_n = 0$.

Every $L$-twisted endomorphism $m : E → E ⊗ L$ has an associated $L$-twisted characteristic polynomial which is a global section $p_m(x) ∈ H^0(\mathbb{L}, (π_{\mathbb{L}}^* L)^\otimes n)$. A global version of the Cayley–Hamilton theorem says that $m$ automatically satisfies its $L$-twisted characteristic polynomial. In particular, if the zero set of $p_m(x)$ is $D$ then $E$ can naturally be thought of as a $π_\pi O_D$-module. Fixing $s ∈ H^0(\mathbb{L}, (π_{\mathbb{L}}^* L)^\otimes n)$, which cuts out the integral curve $D$, [Beauville et al. 1989, Proposition 3.6] gives
the beautiful correspondence
\[ \{ E \to E \otimes L \mid E \text{ a vector bundle and } p_m(x) = s \} \]
\[ \leftrightarrow \{ \text{invertible sheaves } \mathcal{M} \text{ on } D \}. \]  \(\heartsuit\)

The correspondence going from right to left is given by taking the coordinate section of \( \pi^* L \), restricting to \( D \), twisting by \( \mathcal{M} \), and pushing forward along \( \pi \).

To prove Theorem D we need the following key lemma, which provides sufficient conditions for when a section of \( \text{End}(E) \otimes L \) produces a smooth spectral curve.

**Key Lemma.** If \( C \) is a smooth connected genus-\( g \) curve, \( E \) is a rank-\( n \) semistable vector bundle on \( C \), and \( L \) is an ample line bundle on \( C \) with \( \deg(L) \geq 2g \), then the spectral curve associated to a generic section of \( \text{End}(E) \otimes L \) is smooth and irreducible.

The method of proof of the Key Lemma involves a standard analysis of the discriminant locus, where a section of \( \text{End}(E) \otimes L \) has eigenvalues with multiplicity \( \geq 2 \). Before proving the Key Lemma, we show that Theorem D follows immediately.

**Proof of Theorem D.** Let \( C \) be a smooth projective genus-\( g \) curve and \( E \) a rank-\( n \) semistable vector bundle on \( C \). Let \( L \) be a line bundle on \( C \) of degree \( \geq 2g \). By the Key Lemma, if
\[ m : E \to E \otimes L \]
is a general \( L \)-twisted endomorphism then the resulting \( L \)-twisted characteristic polynomial is smooth and irreducible.

Thus, by the correspondence (\( \heartsuit \)) there is a line bundle \( \mathcal{M} \) on \( D \) such that \( \pi_* \mathcal{M} \cong E \). The genus of \( D \) is \( g_D = \binom{r}{2} \deg(L) + n(g - 1) + 1 \) and is independent of \( E \). However, the degree of \( \mathcal{M} \) is \( \deg(E) + \binom{r}{2} \deg(L) \) and does depend on the degree of \( E \). In particular, if
\[ \deg(E) \geq \binom{r}{2} \deg(L) + r(2g - 2) + 3 \]
then \( \mathcal{M} \) is very ample and three general sections of \( \mathcal{M} \) give a map \( \phi : D \to \mathbb{P}^2 \) such that the induced maps \( \pi \times \phi : D \to C \times \mathbb{P}^2 \) and \( \psi_{\pi,\phi} : C \to (\mathbb{P}^2)^{[n]} \) are embeddings. Under the embedding \( \psi_{\pi,\phi} \), the restriction of \( \mathcal{O}_{\mathbb{P}^2}(1)^{[n]} \) to \( C \) is precisely \( E \), proving Theorem D.

We now proceed with the proof of the Key Lemma.

**Lemma 3.1.** If a subvariety \( X \subset E \) of a globally generated vector bundle \( E \) over a smooth curve \( C \) has codimension \( \geq 2 \) then a generic section of \( E \) avoids \( X \). If \( X \subset E \) is a reduced divisor then a generic section of \( E \) meets \( X \) transversely.
Proof. This is an elementary dimension count using generic smoothness in characteristic 0 and the incidence correspondence

\[ I = \{(w, e_x, x) \in W \times \mathbb{E}|_x \times C \mid w(x) = e_x\} \subset W \times \mathbb{E}, \]

where \( W \) is a subspace of sections of \( \mathbb{E} \rightarrow C \) that globally generate \( \mathbb{E} \). The key point is that the projection from \( I \) to \( \mathbb{E} \) is an affine bundle, so the total space of \( I \) is smooth. \( \square \)

If \( \mathbb{H} \) is the total space of \( \mathcal{E}\text{nd}(\mathcal{E}) \otimes L \), and \( \mathcal{C} = L \oplus \cdots \oplus L^\otimes n \), then there is a map \( \epsilon : \mathbb{H} \rightarrow \mathcal{C} \) which sends an \( L \)-twisted endomorphism to the coefficients of its characteristic polynomial. There is a reduced and irreducible divisor in \( \mathbb{U} \subset \mathcal{C} \) which consists of characteristic polynomials with multiple roots. Let \( \mathbb{V} \subset \mathbb{H} \) be the scheme-theoretic inverse of \( \mathbb{U} \).

**Lemma 3.2.** \( \mathbb{V} \) is reduced and irreducible. If a section \( s : C \rightarrow \mathbb{H} \) meets \( \mathbb{V} \) transversely and avoids the locus in \( \mathbb{V} \) with more than one repeated eigenvalue or an eigenvalue of multiplicity \( \geq 3 \), then the corresponding spectral curve is smooth.

**Proof.** First, local trivialization of \( \mathbb{H}, \mathbb{U}, \mathbb{V} \) and \( L \) implies it is enough to check on a fiber. Over a point \( x \in C \) we have \( \mathbb{H}|_x \cong \text{Mat}_{n \times n}(k) \) and \( \mathcal{C}|_x \cong \mathbb{A}^n \). Let \( \mathbb{V}|_x \) be the locus of matrices whose eigenvalues have multiplicity \( \geq 2 \), and let \( \mathbb{U}|_x \) be the discriminant locus. Irreducibility of \( \mathbb{V}|_x \) follows from [Arnold 1971, §5.6], and the fact that it is reduced follows from the observation that \( d\epsilon|_{x,M} \) has maximal rank for a general matrix \( M \in \mathbb{U}|_x \). For the last statement in the lemma, it suffices to verify smoothness for an eigenvalues cover associated to a 1-dimensional family of matrices which meets the discriminant locus transversely at matrices with exactly one repeated eigenvalue; this is a straightforward local calculation. \( \square \)

**Proof of Key Lemma.** Semistability of \( \mathcal{E} \) and the inequality \( \deg L \geq 2g \) imply that \( \mathcal{E}\text{nd}(\mathcal{E}) \otimes L \) is globally generated. By Lemma 3.1 and the first part of Lemma 3.2, a generic section \( s \) of \( \mathcal{E}\text{nd}(\mathcal{E}) \otimes L \) meets \( \mathbb{V} \) transversely and avoids the locus with more than one repeated eigenvalue or an eigenvalue of multiplicity of \( \geq 3 \). By the second part of Lemma 3.2, the associated spectral curve is smooth. By construction of the spectral curve \( C_s \) we have

\[ \pi_* \mathcal{O}_{C_s} \cong \mathcal{O}_C \oplus \cdots \oplus L^{-(n-1)}. \]

Since we assumed \( L \) is ample, \( H^0(C, \pi_* \mathcal{O}_{C_s}) = H^0(C, \mathcal{O}_C) = H^0(C, \mathcal{O}_C) \) is 1-dimensional. Thus \( C_s \) is connected and smooth, so it is irreducible. \( \square \)

### 4. Perturbation of polarization and stability

The goal of this section is to prove (in Proposition 4.7) that the pullback of a stable bundle to a product is stable with respect to a product polarization. Proposition 4.7
was important in the proof of Theorem A. We also prove that stability of the tautological bundles with respect to the natural Chow divisors implies stability with respect to nearby ample divisors. Our approach to proving both of these facts involves considering stability with respect to numerical classes of curves so that we can apply ideas of convexity. In particular, our approach follows ideas appearing recently in [Greb and Toma 2013; Greb et al. 2016] and we recommend looking at these articles to see how these ideas can be developed further and systematically.

Throughout this section, denote by $X$ a normal complex projective variety of dimension $d$. Let $\gamma \in N_1(X)_{\mathbb{R}}$ be a real curve class and let $E$ be a torsion-free sheaf on $X$. For any sheaf $Q$ on $X$, we denote by $\text{Sing}(Q)$ the closed locus where $Q$ is not locally free.

**Definition 4.1.** The slope of $E$ with respect to $\gamma$ is the real number

$$\mu^{\gamma}(E) := \frac{c_1(E) \cdot \gamma}{\text{rank}(E)}.$$  

**Remark 4.2.** Fixing an ample class $H \in N_1(X)_{\mathbb{R}}$, it is true that $\mu^{H}(E) = \mu^{H(d-1)}(E)$. Nonetheless, to distinguish the concepts we use subscripts to denote slope with respect to an ample divisor and superscripts to denote slope with respect to a curve class.

**Definition 4.3.** We say $E$ is slope-stable (resp. slope-semistable) with respect to $\gamma$ if, for all torsion-free quotients $E \to Q \to 0$ of intermediate rank, we have

$$\mu^{\gamma}(E) < \mu^{\gamma}(Q) \quad (\text{resp. } \mu^{\gamma}(E) \leq \mu^{\gamma}(Q)).$$

A benefit of working with slope-(semi)stability with respect to curves rather than divisors is that we can apply ideas of convexity.

**Lemma 4.4.** If $\gamma, \delta$ are classes in $N_1(X)_{\mathbb{R}}$ such that $E$ is semistable with respect to $\gamma$ and $E$ is stable with respect to $\delta$, then $E$ is stable with respect to $a\gamma + b\delta$ for $a, b > 0$. \hfill $\square$

If $C \subset X$ is an irreducible curve, we would like to relate the stability of $E|_C$ and the stability of $E$ with respect to the class of $C$. However, if $Q$ is a coherent sheaf and $C$ meets $\text{Sing}(Q)$, it is possible that $c_1(Q|_C) \neq c_1(Q)|_C$. Thankfully we can say something if $C$ is not entirely contained in $\text{Sing}(Q)$.

**Proposition 4.5.** Let $E \to Q \to 0$ be a torsion-free quotient which destabilizes $E$ with respect to the curve class $\gamma$. Suppose $C \subset X$ is a smooth irreducible closed curve which represents $\gamma$, avoids $\text{Sing}(E)$, and avoids the singularities of $X$. If $C$ is not contained in $\text{Sing}(Q)$ then $E|_C$ is not stable on $C$.  


Proof. First, we can reduce to the surface case by choosing a normal surface $S \subset X$ containing $C$ such that $S$ is smooth along $C$, and $S$ meets Sing($Q$) and Sing($E$) properly. This is possible because when the dimension of $X$ is greater than 3 a generic, high-degree hyperplane section containing $C$ is normal and smooth along $C$ and meets both Sing($Q$) and Sing($E$) properly. Once such a surface is chosen, we have

$$c_1(Q)|_S = c_1(Q)|_S = c_1(Q|_S / \text{Tors}(Q|_S)),$$

because both Sing($Q$) $\cap$ $S$ and Sing($E$) $\cap$ $S$ are zero-dimensional. Thus

$$E|_S / \text{Tors}(E|_S) \rightarrow Q|_S / \text{Tors}(Q|_S) \rightarrow 0$$

is a torsion-free quotient on $S$ which destabilizes $E|_S / \text{Tors}(E|_S)$ with respect to the class of $C$. So we have reduced the proposition to the case when $X$ is a surface.

Let $X$ be a surface. It is enough to show $c_1(Q|_C) = c_1(Q)|_C$. The restriction $c_1(Q)|_C$ is computed via the derived pullback

$$c_1(Q)|_C = \sum_{i=0}^{\infty} (-1)^i c_1(\text{Tor}_i^{O_X}(Q, O_C)),$$

where the Tor$_i^{O_X}(Q, O_C)$ are thought of as modules on $C$ (see [Fulton 1998, §15.1] for the smooth case). Further, $C$ is a Cartier divisor on $X$, so $O_C$ has a two-term locally free resolution. So the Tor$_i^{O_X}(Q, O_C)$ vanish for $i > 2$ and Tor$_1^{O_X}(Q, O_C) = 0$ because $Q$ is torsion-free. Therefore,

$$c_1(Q)|_C = c_1(\text{Tor}_0^{O_X}(Q, O_C)) = c_1(Q|_C).$$

So $E|_C$ is not slope-stable. □

An immediate corollary is the following coarse criterion for checking slope-stability with respect to $\gamma$.

**Corollary 4.6.** Let $\pi : C_T \rightarrow T$ be a family of smooth irreducible closed curves in $X$ with class $\gamma$. For $t \in T$ we write $C_t$ to denote $\pi^{-1}(t)$. Suppose $E$ is a vector bundle on $X$ such that $E|_{C_t}$ is stable for all $t \in T$. If the curves in $C_T$ are dense in $X$ then $E$ is stable with respect to the curve class $\gamma$.

**Proof.** Suppose for contradiction that $E$ is unstable with respect to $\gamma$. Then there exists a torsion-free quotient $E \rightarrow Q \rightarrow 0$ with $\mu^\gamma(Q) \leq \mu^\gamma(E)$. As $Q$ is torsion-free, Sing($Q$) has codimension $\geq 2$. The curves in $C_T$ are dense in $X$ so there is a $t \in T$ such that $C_t$ is not contained in Sing($Q$). Then Proposition 4.5 guarantees that $E|_{C_t}$ is not stable, which contradicts our hypothesis. □

Proposition 4.5 can be adjusted so that Corollary 4.6 also holds if stability is replaced by semistability. As a consequence we prove the following basic result
about slope-stable vector bundles, which we have already used in the proof of Theorem A.

**Proposition 4.7.** Let $X$ and $Y$ be smooth projective varieties of dimension $d$ and $e$, respectively. Let $H_X$ be an ample divisor on $X$ and let $H_Y$ be an ample divisor on $Y$. Let $p_1$ denote the projection from $X \times Y$ to $X$ and $p_2$ the projection from $X \times Y$ to $Y$. If $\mathcal{E}$ is a vector bundle on $X$ which is slope-stable with respect to $H_X$, then $p_1^*(\mathcal{E})$ is slope-stable on $X \times Y$ with respect to the ample divisor $p_1^*(H_X) + p_2^*(H_Y)$.

**Proof.** By [Mehta and Ramanathan 1984, Theorem 4.3] if $k \gg 0$ and $C$ is a general curve which is a complete intersection of divisors linearly equivalent to $kH_X$ then $\mathcal{E}|_C$ is stable. Let $F \subset |kH_X|^{d-1}$ be the open subset of the cartesian power of the complete linear series of $kH_X$ defined as
\[ F := \{(H_1, \ldots, H_{d-1}) \in |kH_X|^{d-1} \mid C = H_1 \cap \cdots \cap H_{d-1} \} \]
is a smooth complete intersection curve and $\mathcal{E}|_C$ is stable.

We write $C_F$ for the natural family of smooth curves in $X$ parametrized by $F$. Likewise, the fiber product $C_F \times_F (F \times Y)$ is naturally a family of smooth curves in $X \times Y$ parametrized by $F \times Y$. The image of $C_F \times_F (F \times Y)$ in $X \times Y$ is dense, and for any $(f, y) \in F \times Y$ the restriction of $p_1^*(\mathcal{E})$ to $C_{(f,y)}$ is stable. Therefore, by Corollary 4.6, $p_1^*(\mathcal{E})$ is stable with respect to the numerical class of $C_{(f,y)}$, which we denote by $\gamma$.

For $l \gg 0$ the divisor $lH_Y$ is very ample on $Y$ and a general complete intersection of divisors linearly equivalent to $lH_Y$ is smooth. Let $G \subset |lH_Y|^{e-1}$ be the open subset of the cartesian power of the complete linear series of $lH_Y$ defined as
\[ G := \{(H_1, \ldots, H_{e-1}) \in |lH_Y|^{e-1} \mid H_1 \cap \cdots \cap H_{e-1} \} \]
is a smooth complete intersection curve.

As before, there is a natural family $D_G$ of smooth curves in $Y$ parametrized by $G$. The fiber product $D_G \times_G (X \times G)$ is a family of smooth curves in $X \times Y$ parametrized by $X \times G$. For $(x, g) \in X \times G$ the restriction of $p_1^*(\mathcal{E})$ to $D_{(x,g)}$ is a direct sum of trivial bundles, thus the restriction is semistable. Therefore, by applying Corollary 4.6 in the semistable case, $p_1^*(\mathcal{E})$ is semistable with respect to the curve class of $D_{(x,g)}$, which we denote by $\delta$.

Finally,
\[(p_1^*H_X + p_2^*H_Y)^{d+e-1} = \left( \frac{d+e-1}{e} \right) \left( \frac{H_Y}{k^{d-1}} \right)^{\gamma \cdot l} + \left( \frac{d+e-1}{d} \right) \left( H_X \right)^d \left( \frac{l}{e-1} \right) . \delta. \]
Thus, by Lemma 4.4, $p_1^*(\mathcal{E})$ is slope-stable with respect to $p_1^*(H_X) + p_2^*(H_Y)$. □

This completes the proof of Theorem A. We now give a proof of the perturbation argument. The idea is to use [Greb et al. 2016, Theorem 3.4] on openness of...
stability along with the fact that the natural Chow divisors are nef in the sense of [de Cataldo and Migliorini 2002, Definition 2.1.3].

**Proposition 4.8.** Let $H$ be a nef divisor and $A$ an ample $\mathbb{Q}$-divisor on a normal complex projective variety $X$. Suppose $E$ is a rank-$r$ torsion-free sheaf on $X$ which is slope-stable with respect to the class of $H^{d-1}$. Assume

$$- \cap H^{d-2} : N^1(X)_\mathbb{R} \to N_1(X)_\mathbb{R}, \quad \xi \mapsto \xi \cdot H^{d-2}$$

is an isomorphism. Then $E$ is stable with respect to $H + \epsilon A$ for $\epsilon$ sufficiently small.

This implies that we can perturb our Chow polarization to obtain stability of tautological bundles with respect to nearby ample divisors.

**Corollary 4.9.** If $E$ is a vector bundle on a smooth projective surface $S$ which is stable with respect to an ample divisor $H$, then $E^{[n]}$ is stable with respect to an ample divisor near the Chow divisor $H_n$.

**Proof of Corollary 4.9.** By [de Cataldo and Migliorini 2002, Theorem 2.3.1] we know $H_n$ is lef, so $E^{[n]}$ and $H_n$ satisfy the conditions of Proposition 4.8. Therefore, $E^{[n]}$ is stable with respect to ample divisors close to $H_n$. \qed

**Proof of Proposition 4.8.** Identifying the tangent space of a vector space with the vector space, the derivative of the $(d-1)$-st power map $N^1(X)_\mathbb{R} \to N_1(X)_\mathbb{R}$ at $H$ is given by

$$- \cap (d-1)H^{d-2} : N^1(X)_\mathbb{R} \to N_1(X)_\mathbb{R}.$$ 

The assumption that the intersection with the $H^{d-2}$ map is an isomorphism implies that the $(d-1)$-st power map is locally an isomorphism.

It follows from [Greb et al. 2016, Theorem 3.4] that there is a nonempty convex open set $U \subset N_1(X)_\mathbb{R}$ whose closure contains $[H^{d-1}]$ such that, for all $\gamma \in U$, $E$ is stable with respect to $\gamma$. More precisely, if $\delta \in N_1(X)_\mathbb{R}$ represents the $(d-1)$-st power of an ample divisor then $E$ is stable with respect to the perturbed curve class $[H^{d-1}] + \epsilon \cdot \delta$ for $\epsilon$ sufficiently small. By estimating the $(d-1)$-st power map by its derivative (which is an isomorphism at $H$) and by our ability to perturb linearly towards ample curve classes, we see that, for small enough $\epsilon$, $(H + \epsilon A)^{d-1}$ maps into $U$. Therefore, for $\epsilon$ sufficiently small, $E$ is stable with respect to $H + \epsilon A$. \qed

**Acknowledgements**

I am grateful to my advisor, Robert Lazarsfeld, who suggested the project and directed me in productive lines of thought. I am also thankful for conversations and correspondence with Lawrence Ein, Roman Gayduk, Daniel Greb, Julius Ross, Giulia Saccà, Ian Shipman, Brooke Ullery, Dingxin Zhang, and Xin Zhang. This paper is a substantial revision of a previous preprint. I would finally like to thank the referees for thoroughly reviewing the paper and offering helpful suggestions.
References


Communicated by David Eisenbud
Received 2015-06-28 Revised 2016-04-28 Accepted 2016-05-28
david.stapleton@stonybrook.edu Department of Mathematics, Stony Brook University, Math Tower 2118, Stony Brook, NY 11794, United States
Anabelian geometry
and descent obstructions on moduli spaces

Stefan Patrikis, José Felipe Voloch and Yuri G. Zarhin

We study the section conjecture of anabelian geometry and the sufficiency of
the finite descent obstruction to the Hasse principle for the moduli spaces of
principally polarized abelian varieties and of curves over number fields. For the
former we show that the section conjecture fails and the finite descent obstruction
holds for a general class of adelic points, assuming several well-known conjec-
tures. This is done by relating the problem to a local-global principle for Galois
representations. For the latter, we show how the sufficiency of the finite descent
obstruction implies the same for all hyperbolic curves.

1. Introduction

Anabelian geometry is a program proposed by Grothendieck [1997a; 1997b] which
suggests that for a certain class of varieties (called anabelian but, as yet, undefined)
over a number field, one can recover the varieties from their étale fundamental
group together with the Galois action of the absolute Galois group of the number
field. Precise conjectures exist only for curves and some of them have been proved,
notably by Mochizuki [1996]. Grothendieck suggested that moduli spaces of curves
and abelian varieties (the latter perhaps less emphatically) should be anabelian.
Already Ihara and Nakamura [1997] have shown that moduli spaces of abelian
varieties should not be anabelian as one cannot recover their automorphism group
from the fundamental group and we will further show that other anabelian properties
fail in this case.

The finite descent obstruction is a construction that describes a subset of the
adelic points of a variety over a number field containing the closure of the rational
(or integral) points and is conjectured, for hyperbolic curves (Stoll [2007] in the
projective case and Harari and Voloch [2010] in the affine case), to equal that
closure. It’s not unreasonable to conjecture the same for all anabelian varieties.
The relationship between the finite descent obstruction and the section conjecture
in anabelian geometry has been discussed by Harari and Stix [2012], Stix [2013,

MSC2010: primary 11G35; secondary 14G05, 14G35.
Keywords: Anabelian geometry, moduli spaces, abelian varieties, descent obstruction.
Section 11), and others. We will review the relevant definitions below, although our point of view will be slightly different.

The purpose of this paper is to study the section conjecture of anabelian geometry and the finite descent obstruction for the moduli spaces of principally polarized abelian varieties and of curves over number fields. For the moduli of abelian varieties we show that the section conjecture fails in general and that both the section conjecture and finite descent obstruction hold for a general class of adelic points, assuming many established conjectures in arithmetic geometry (specifically, we assume the Hodge, Tate, Fontaine–Mazur and Grothendieck–Serre conjectures, in the precise forms stated in Section 3). This is done by converting the question into one about Galois representations.

The section conjecture predicts that sections of the fundamental exact sequence (Section 3, Equation (1)) of an anabelian variety over a number field correspond to rational points. In this paper, we look at the sections of the fundamental exact sequence of the moduli spaces of principally polarized abelian varieties that, locally at every place of the ground field, come from a point rational over the completion, which moreover is integral for all but finitely many places. This set is denoted $S_0(K, A_g)$ and defined precisely at the end of Section 2. We explain, in Section 3, how sections of the fundamental exact sequence of the moduli spaces of principally polarized abelian varieties correspond to Galois representations and prove, Theorem 3.7, the following result.

**Theorem 1.1.** Assume the Hodge, Tate, Fontaine–Mazur, and Grothendieck–Serre conjectures. Let $K$ be a number field. Suppose $s \in S_0(K, A_g)$ gives rise to a system of $\ell$-adic Galois representations one of which is absolutely irreducible. Then there exists, up to isomorphism, a unique principally polarized abelian variety which, viewed as point of $A_g(K)$, induces (up to conjugation) the section $s$.

We also give examples (see Theorems 4.4 and 4.5) showing that weaker versions of the above result do not hold. Specifically, the local conditions cannot be weakened to hold almost everywhere, for instance.

For the moduli of curves, we show how combining some of our results and assuming sufficiency of finite descent obstruction for the moduli of curves, we deduce the sufficiency of finite descent obstruction for all hyperbolic curves.

In the next section we give more precise definitions of the objects we use and in the following two sections we give the applications mentioned above.

### 2. Preliminaries

Let $X/K$ be a smooth geometrically connected variety over a field $K$. Let $G_K$ be the absolute Galois group of $K$ and $\bar{X}$ the base-change of $X$ to an algebraic closure of $K$. We denote by $\pi_1(\cdot)$ the algebraic fundamental group functor on
(geometrically pointed) schemes and we omit base-points from the notation. We have the fundamental exact sequence
\[ 1 \to \pi_1(\bar{X}) \to \pi_1(X) \to G_K \to 1. \] (1)

The map \( p_X : \pi_1(X) \to G_K \) from the above sequence is obtained by functoriality from the structural morphism \( X \to \text{Spec } K \). Grothendieck’s anabelian program is to specify a class of varieties, termed anabelian, for which the varieties and morphisms between them can be recovered from the corresponding fundamental groups together with the corresponding maps \( p_X \) when the ground field is finitely generated over \( \mathbb{Q} \). As this is very vague, we single out here two special cases with precise statements. The first is a (special case of a) theorem of Mochizuki [1996] which implies part of Grothendieck’s conjectures for curves but also extends it by considering \( p \)-adic fields.

**Theorem 2.1** [Mochizuki 1996]. Let \( X, Y \) be smooth projective curves of genus bigger than one over a field \( K \) which is a subfield of a finitely generated extension of \( \mathbb{Q}_p \). If there is an isomorphism from \( \pi_1(X) \) to \( \pi_1(Y) \) inducing the identity on \( G_K \) via \( p_X, p_Y \), then \( X \) is isomorphic to \( Y \).

A point \( P \in X(K) \) gives, by functoriality, a section \( G_K \to \pi_1(X) \) of the fundamental exact sequence (1) well-defined up to conjugation by an element of \( \pi_1(\bar{X}) \) (the indeterminacy is because of base points).

We denote by \( H(K, X) \) the set of sections \( G_K \to \pi_1(X) \) modulo conjugation by \( \pi_1(\bar{X}) \) and we denote by \( \sigma_{X/K} : X(K) \to H(K, X) \) the map that associates to a point the class of its corresponding section, as above, and we call it the section map. As part of the anabelian program, it is expected that \( \sigma_{X/K} \) is a bijection if \( X \) is projective, anabelian and \( K \) is finitely generated over its prime field. This is widely believed in the case of hyperbolic curves over number fields and is usually referred as the section conjecture. For a similar statement in the nonprojective case, one needs to consider the so-called cuspidal sections, see [Stix 2013, Section 18]. Although we will discuss nonprojective varieties in what follows, we will not need to specify the notion of cuspidal sections. The reason for this is that we will be considering sections that locally come from points (the Selmer set defined below) and these will not be cuspidal.

We remark that the choice of a particular section \( s_0 : G_K \to \pi_1(X) \) induces an action of \( G_K \) on \( \pi_1(\bar{X}) \), \( x \mapsto s_0(\gamma)x s_0(\gamma)^{-1} \). For an arbitrary section \( s : G_K \to \pi_1(X) \) the map \( \gamma \mapsto s(\gamma)s_0(\gamma)^{-1} \) is a 1-cocycle for the above action of \( G_K \) on \( \pi_1(\bar{X}) \) and this induces a bijection \( H^1(G_K, \pi_1(\bar{X})) \to H(K, X) \). We stress that this only holds when \( H(K, X) \) is nonempty and a choice of \( s_0 \) can be made. It is possible for \( H(K, X) \) to be empty, in which case there is no natural choice of action of \( G_K \)
on $\pi_1(\overline{X})$ by which to define $H^1(G_K, \pi_1(\overline{X}))$, which would be nonempty in any case, if defined.

Let $X/K$ be as above, where $K$ is now a number field. If $v$ is a place of $K$, we have the completion $K_v$, and a fixed inclusion $\overline{K} \subset K_v$ induces a map $\alpha_v : G_K \to G_K$ and a map $\beta_v : \pi_1(X_v) \to \pi_1(X)$, where $X_v$ is the base-change of $X$ to $K_v$. We define the Selmer set of $X/K$ as the set $S(K, X) \subset H(K, X)$ consisting of the equivalence classes of sections $s$ such that for all places $v$ there exists $P_v \in X(K_v)$ with $s \circ \alpha_v = \beta_v \circ \sigma_{X_v/K_v}(P_v)$. Note that if $v$ is complex, then the condition at $v$ is vacuous and that if $v$ is real, $\sigma_{X_v/K_v}$ factors through $X(K_v)_e$, the set of connected components of $X(K_v)$, equipped with the quotient topology (see [Mochizuki 2003; Pál 2011]). In the nonarchimedian case, $X(K_v)$ is totally disconnected so $X(K_v) = X(K_v)_e$, and we have the following diagram:

$$\begin{array}{ccc}
X(K) & \longrightarrow & \prod v X(K_v)_e \supset X^f \\
\sigma_{X/K} & \downarrow & \downarrow \prod \sigma_{X_v/K_v} \\
S(K, X) \subset H(K, X) & \longrightarrow & \prod \alpha H(K_v, X_v)
\end{array}$$

We define the set $X^f$ (the finite descent obstruction) as the set of points $(P_v)_v \in \prod_v X(K_v)_e$, for which there exists $s \in H(K, X)$ (which is then necessarily an element of $S(K, X)$) satisfying $s \circ \alpha_v = \beta_v \circ \sigma_{X_v/K_v}(P_v)$ for all places $v$. Also, it is clear that the image of $X(K)$ is contained in $X^f$. At least when $X$ is proper, $X^f$ is closed (this follows from the compactness of $H(K, X)$ [Stix 2013, Corollary 45]). In that case, one may consider whether the closure of the image of $X(K)$ in $\prod_v X(K_v)_e$ equals $X^f$. A related statement is the equality $\sigma_{X/K}(X(K)) = S(K, X)$, which is implied by the “section conjecture”, i.e., the bijectivity of $\sigma_{X/K} : X(K) \to H(K, X)$.

As a specific instance of this relation, we record the following easy fact.

**Proposition 2.2.** We have that $X^f = \emptyset$ if and only if $S(K, X) = \emptyset$.

**Proof.** If $X^f \neq \emptyset$ and $(P_v) \in X^f$, then there exists $s \in S(K, X)$ with $s \circ \alpha_v = \beta_v \circ \sigma_{X_v/K_v}(P_v)$ for all places $v$, so $S(K, X) \neq \emptyset$.

If $s \in S(K, X)$, there exists $(P_v)$ with $s \circ \alpha_v = \beta_v \circ \sigma_{X_v/K_v}(P_v)$ for all places $v$. So $(P_v) \in X^f$. $\square$

If $X$ is not projective, then one has to take into account questions of integrality. We choose an integral model $\mathcal{X}/\mathcal{O}_{S,K}$, where $S$ is a finite set of places of $K$ and $\mathcal{O}_{S,K}$ is the ring of $S$-integers of $K$. The image of $X(K)$ in $X^f$ actually lands in the adelic points which are the points that satisfy $P_v \in \mathcal{X}(\mathcal{O}_v)$ for all but finitely many $v$, where $\mathcal{O}_v$ is the local ring at $v$. Similarly, the image of $\sigma_{X/K}$ belongs to the subset of $S(K, X)$ where the corresponding local points $P_v$ also belong to $\mathcal{X}(\mathcal{O}_v)$ for all but finitely many $v$. We denote this subset of $S(K, X)$ by $S_0(K, X)$ and call
it the integral Selmer set. We note that $S_0(K, X)$ is independent of the choice of the model $\mathcal{X}$.

In order to set notation, we recall here some basic notions about the Tate module of abelian varieties which will be used in the next two sections. If $A$ is an abelian variety over the field $K$ then we write $\text{End}(A)$ for its ring of all $K$-endomorphisms and $\text{End}^0(A)$ for the corresponding (finite-dimensional semisimple) $\mathbb{Q}$-algebra $\text{End}(A) \otimes \mathbb{Q}$. If $n \geq 3$ is an integer that is not divisible by $\text{char}(K)$ and all points of order $n$ on $A$ are defined over $K$ then, by a theorem of Silverberg [1992], all $K$-endomorphisms of $A$ are defined over $K$, i.e., lie in $\text{End}(A)$.

If $\ell$ is a prime different from $\text{char}(K)$ then we write $T_\ell(A)$ for the $\mathbb{Z}_\ell$-Tate module of $A$ which is a free $\mathbb{Z}_\ell$-module of rank $2 \dim(A)$ provided with the natural continuous homomorphism $\rho_{\ell, A} : G_K \to \text{Aut}_{\mathbb{Z}_\ell}(T_\ell(A))$

and the $\mathbb{Z}_\ell$-ring embedding

$$e_\ell : \text{End}(A) \otimes \mathbb{Z}_\ell \hookrightarrow \text{End}_{\mathbb{Z}_\ell}(T_\ell(A)).$$

The image of $\text{End}(A) \otimes \mathbb{Z}_\ell$ commutes with $\rho_{\ell, A}(G_K)$. Tensoring by $\mathbb{Q}_\ell$ (over $\mathbb{Z}_\ell$), we obtain the $\mathbb{Q}_\ell$-Tate module of $A$

$$V_\ell(A) = T_\ell(A) \otimes_{\mathbb{Z}_\ell} \mathbb{Q}_\ell,$$

which is a $2 \dim(A)$-dimensional $\mathbb{Q}_\ell$-vector space containing

$$T_\ell(A) = T_\ell(A) \otimes 1$$

as a $\mathbb{Z}_\ell$-lattice. We may view $\rho_{\ell, A}$ as an $\ell$-adic representation

$$\rho_{\ell, A} : G_K \to \text{Aut}_{\mathbb{Z}_\ell}(T_\ell(A)) \subset \text{Aut}_{\mathbb{Q}_\ell}(V_\ell(A))$$

and extend $e_\ell$ by $\mathbb{Q}_\ell$-linearity to the embedding of $\mathbb{Q}_\ell$-algebras

$$\text{End}^0(A) \otimes_{\mathbb{Q}} \mathbb{Q}_\ell = \text{End}(A) \otimes \mathbb{Q}_\ell \hookrightarrow \text{End}_{\mathbb{Q}_\ell}(V_\ell(A)),$$

which we still denote by $e_\ell$. Further we will identify $\text{End}^0(A) \otimes_{\mathbb{Q}} \mathbb{Q}_\ell$ with its image in $\text{End}_{\mathbb{Q}_\ell}(V_\ell(A))$.

This provides $V_\ell(A)$ with the natural structure of $G_K$-module; in addition, $\text{End}^0(A) \otimes_{\mathbb{Q}} \mathbb{Q}_\ell$ is a $\mathbb{Q}_\ell$-(sub)algebra of endomorphisms of the Galois module $V_\ell(A)$. In other words,

$$\text{End}^0(A) \otimes_{\mathbb{Q}} \mathbb{Q}_\ell \subset \text{End}_{G_K}(V_\ell(A)).$$

Let $\chi_\ell$ be the cyclotomic character $\chi_\ell : G_K \to \mathbb{Z}_\ell^*$ that defines the Galois action on all $\ell$-power roots of unity, and $\mathbb{Z}_\ell(1)$ the $\ell$-adic Tate module of the multiplicative
group $G_m$. The group $\mathbb{Z}_\ell(1)$ is a free $\mathbb{Z}_\ell$-module of rank 1 provided with the Galois action that is defined by

$$\chi_\ell : G_K \to \mathbb{Z}_\ell^* = \text{Aut}_{\mathbb{Z}_\ell}(\mathbb{Z}_\ell(1)).$$

Let $\hat{A}$ be the dual (Picard) variety of $A$ [Lang 1959; Mumford 1970], which is an abelian variety over $K$ that is isogenous to $A$. There is the Weil pairing [Lang 1959, Chapter VII, Section 2]

$$e_\ell : T_\ell(A) \times T_\ell(\hat{A}) \to \mathbb{Z}_\ell(1),$$

which is a Galois-equivariant, $\mathbb{Z}_\ell$-bilinear perfect/unimodular pairing of free $\mathbb{Z}_\ell$-modules $T_\ell(A)$ and $T_\ell(\hat{A})$. This implies that the Galois modules $T_\ell(\hat{A})$ and $\text{Hom}_{\mathbb{Z}_\ell}(T_\ell(A), \mathbb{Z}_\ell(1))$ are isomorphic.

### 3. Moduli of abelian varieties

The moduli space of principally polarized abelian varieties of dimension $g$ is denoted by $A_g$. It is actually a Deligne–Mumford stack or orbifold and we will consider its fundamental group as such. For a general definition of fundamental groups of stacks including a proof of the fundamental exact sequence in this generality, see Zoonekynd 2001. For a discussion of the case of $A_g$, see Hain 2011. We can also get what we need from Ihara and Nakamura 1997 (see below) or by working with a level structure which brings us back to the case of smooth varieties.

As $A_g$ is defined over $\mathbb{Q}$, we can consider it over an arbitrary number field $K$. As per our earlier conventions, $\overline{A}_g$ is the base change of $A_g$ to an algebraic closure of $\mathbb{Q}$ and not a compactification. In fact, we will not consider a compactification at all here. The topological fundamental group of $\overline{A}_g$ is the symplectic group $\text{Sp}_{2g}(\mathbb{Z})$ and the algebraic fundamental group is its profinite completion. When $g > 1$ (which we henceforth assume) $\text{Sp}_{2g}(\mathbb{Z})$ has the congruence subgroup property [Bass et al. 1964; Mennicke 1965] and therefore its profinite completion is $\text{Sp}_{2g}(\hat{\mathbb{Z}})$.

The group $\pi_1(A_g)$ is essentially described by the exact sequences (3.2) and (3.3) of Ihara and Nakamura 1997 and it follows that the set $H(K, A_g)$ consists of $\hat{\mathbb{Z}}$ representations of $G_K$ of rank $2g$ preserving the symplectic form up to a multiplier given by the cyclotomic character. Indeed, it is clear that every section gives such a representation and the converse follows formally from the diagram below, which is a consequence of (3.2) and (3.3) of Ihara and Nakamura 1997.

In the following we denote the cyclotomic character by $\chi : G_K \to \hat{\mathbb{Z}}^*$.

$$
\begin{array}{ccccccccc}
1 & \longrightarrow & \pi_1(\overline{A}_g) & \longrightarrow & \pi_1(A_g) & \longrightarrow & G_K & \longrightarrow & 1 \\
\downarrow & \cong & \downarrow & & \downarrow & \chi & \downarrow & & \\
1 & \longrightarrow & \text{Sp}_{2g}(\hat{\mathbb{Z}}) & \longrightarrow & \text{GSp}_{2g}(\hat{\mathbb{Z}}) & \longrightarrow & \hat{\mathbb{Z}}^* & \longrightarrow & 1 \\
\end{array}
$$
The coverings of $\tilde{A}_g$ corresponding to the congruence subgroups of $\text{Sp}_{2g}(\hat{\mathbb{Z}})$ are those obtained by adding level structures. In particular, for an abelian variety $A$, $\sigma_{A_g/K}(A) = \prod T_\ell(A)$, the product of its Tate modules considered, as usual, as a $G_K$-module. If $K$ is a number field, whenever two abelian varieties are mapped to the same point by $\sigma_{A_g/K}$, then they are isogenous, by [Faltings 1983]. The finiteness of isogeny classes of polarized abelian varieties over $K$ [Faltings 1983] (see also [Zarhin 1985]) implies that for any given $K$ and $g$ every fiber of $\sigma_{A_g/K}$ is finite. On the other hand, $\sigma_{A_g/K}$ is not necessarily injective to $S_0(K, A_g)$. For example, for each $g$ there exists $K$ with noninjective $\sigma_{A_g/K}$. Regarding surjectivity, we will prove that those elements of $S_0(K, A_g)$ for which the corresponding Galois representation is absolutely irreducible (see below for the precise hypothesis and Theorem 3.7 for a precise statement) are in the image of $\sigma_{A_g/K}$, assuming the Fontaine–Mazur conjecture, the Grothendieck–Serre conjecture on semisimplicity of $\ell$-adic cohomology of smooth projective varieties, and the Tate and Hodge conjectures. The integral Selmer set $S_0(K, A_g)$, defined in the previous section, corresponds to the set of Galois representations that are almost everywhere unramified and, locally, come from abelian varieties (which thus are of good reduction for almost all places of $K$) and we will also consider a few variants of the question of surjectivity of $\sigma_{A_g/K}$ to $S_0(K, A_g)$ by different local hypotheses and discuss what we can and cannot prove. A version of this kind of question has also been considered by B. Mazur [1999].

Here is the setting. Let $K$ be a number field, with $G_K = \text{Gal}(\overline{K}/K)$. Fix a finite set of rational primes $S$, and consider a collection of continuous $\ell$-adic representations $\{\rho_\ell : G_K \to \text{GL}_N(\mathbb{Q}_\ell)\}_{\ell \not\in S}$.

We will say that the collection $\{\rho_\ell\}_{\ell \not\in S}$ is weakly compatible if there exists a finite set of places $\Sigma$ of $K$ such that

1. for all $\ell \not\in S$, $\rho_\ell$ is unramified outside the union of $\Sigma$ and the places $\Sigma_\ell$ of $K$ dividing $\ell$; and
2. for all $v \not\in \Sigma \cup \Sigma_\ell$, denoting by $\text{fr}_v$ a (geometric) frobenius element at $v$, the characteristic polynomial of $\rho_\ell(\text{fr}_v)$ has rational coefficients and is independent of $\ell \not\in S$.\footnote{These systems were introduced by Serre [1989], who called them strictly compatible.}

Our aim is to prove the following:

**Theorem 3.1.** We will assume $\{\rho_\ell\}_{\ell \not\in S}$ is weakly compatible and moreover satisfies the following three conditions:

1. **For some prime $\ell_0 \not\in S$,** $\rho_{\ell_0}$ is de Rham at all places of $K$ above $\ell_0$.
2. **For some prime $\ell_1 \not\in S$,** $\rho_{\ell_1}$ is absolutely irreducible.
For some prime \( \ell_2 \not\in S \), and at least one place \( v|\ell_2 \) of \( K \), \( \rho_{\ell_2}|_{G_{K_v}} \) is de Rham with Hodge–Tate weights \(-1, 0\), each with multiplicity \( N/2 \). (This condition holds if there exists an abelian variety \( A_v/K_v \) such that \( \rho_{\ell_2}|_{G_{K_v}} \cong V_{\ell_2}(A_v) \).)

Assume the Hodge, Tate, Fontaine–Mazur, and Grothendieck–Serre conjectures, and suppose that the set \( S \) is empty. Then there exists an abelian variety \( A \) over \( K \) such that \( \rho_{\ell} \cong V_{\ell}(A) \) for all \( \ell \).

We note that the arguments allow \( \ell_0 = \ell_2 \), and the reader may prefer to think of these together as a single condition; we have phrased it this way to have hypotheses that most clearly match the form of the argument.

We begin by making precise the combined implications of the Grothendieck–Serre, Tate, and Fontaine–Mazur conjectures (the Hodge conjecture will only be used later, in the proof of Lemma 3.5). For any field \( k \) and characteristic zero field \( E \), let \( \mathcal{M}_{k,E} \) denote the category of pure homological motives over \( k \) with coefficients in \( E \) (omitting \( E \) from the notation will mean \( E = \mathbb{Q} \)).

**Lemma 3.2.** Assume the Tate conjecture for all finitely generated extensions \( k \) of \( \mathbb{Q} \). Then:

1. The Lefschetz standard conjecture holds for all fields of characteristic zero.
2. All of the standard conjectures (namely, the Künneth and Hodge standard conjectures, and the agreement of numerical and homological equivalence) hold for all fields of characteristic zero.
3. For any field \( k \) that can be embedded in \( \mathbb{C} \), the category \( \mathcal{M}_k \) is a semisimple neutral Tannakian category over \( \mathbb{Q} \).
4. For any finitely generated \( k/\mathbb{Q} \), the étale \( \ell \)-adic realization functor
   \[
   \mathcal{M}_{k,\mathbb{Q}_\ell} \to \text{Rep}_{\mathbb{Q}_\ell}(G_k),
   \]
   valued in the category of continuous \( \ell \)-adic representations of \( G_k \), is fully faithful.

**Proof.** For the first assertion, see, e.g., [André 2004, 7.3.1.3]; for the second, see [André 2004, 5.4.2.2]. The third part is the basic motivating consequence of the standard conjectures (a fiber functor over \( \mathbb{Q} \) is given by Betti cohomology, after fixing an embedding \( k \hookrightarrow \mathbb{C} \)): see [Jannsen 1992, Corollary 2], especially for the semisimplicity claim. Finally, for the last part, fullness is the Tate conjecture; and faithfulness follows from the agreement of numerical and homological equivalence and [Tate 1994, Lemma 2.5] (note that faithfulness on \( \mathcal{M}_k \) is simply by definition of homological equivalence: it is only with \( \mathbb{Q}_\ell \)-coefficients that some argument is needed). \( \Box \)
For the rest of this section, we assume the Tate conjecture for all finitely generated $k$ of characteristic zero. Thus, we have a motivic Galois formalism: $\mathcal{M}_{k,E}$ is equivalent to $\text{Rep}(G_{k,E})$ for some proreductive group $G_{k,E}$ over $E$, the equivalence depending on the choice of an $E$-linear fiber functor. We will implicitly fix an embedding $k \hookrightarrow \mathbb{C}$ and use the associated Betti realization as our fiber functor. Before proceeding, we introduce two pieces of notation. For an extension of fields $k'/k$, we denote the base-change of motives by

$$(\cdot)|_{k'} : \mathcal{M}_{k,E} \to \mathcal{M}_{k',E}.$$ 

This is not to be confused with the change of coefficients. Fix an embedding $\iota: \mathbb{Q} \hookrightarrow \mathbb{Q}_\ell$, so that when $E$ is a subfield of $\mathbb{Q}$ we can speak of the $\ell$-adic realization $H_\iota : \mathcal{M}_{k,E} \to \text{Rep}_{\mathbb{Q}_\ell}(G_k)$ associated to $\iota$.

Now we turn to the case of number fields, i.e., $k = K$. The Tate conjecture alone does not suffice to link Galois representations with motives: it yields full faithfulness of the $\ell$-adic realization (as in Lemma 3.2), but does not characterize the essential image. This is done via the combination of the Fontaine–Mazur and Grothendieck–Serre semisimplicity conjectures, which we now recall. A semisimple representation $r_\ell : G_K \to \text{GL}_N(\mathbb{Q}_\ell)$ is said to be geometric (in the sense of Fontaine and Mazur [1995]) if it is unramified outside a finite set of places of $K$, and if for all $v | \ell$ of $K$, the restriction $r_\ell|_{G_{K_v}}$ is de Rham (equivalently, potentially semistable, as in the original formulation). See [Fontaine and Ouyang 2007; Brinon and Conrad 2009] for the definition and basic properties of de Rham representations. Fontaine and Mazur have conjectured that any irreducible geometric $r_\ell$ is isomorphic to a subquotient of $H^i(X_{\overline{F}}, \mathbb{Q}_\ell)(j)$ for some smooth projective variety $X/K$ and some integers $i$ and $j$; that the converse assertion holds is a consequence of the base-change theorems of étale cohomology [SGA4 1/2 1977] and the $p$-adic de Rham comparison isomorphism of Faltings [1989]. Grothendieck and Serre have moreover conjectured that for any smooth projective $X/K$, and any integer $i$, $H^i(X_{\overline{F}}, \mathbb{Q}_\ell)$ is a semisimple representation of $G_K$. Putting all of these conjectures together, we can characterize the essential image of $H_\iota$:

**Lemma 3.3.** Assume the Tate, Fontaine–Mazur, and Grothendieck–Serre conjectures. Let $r_\ell : G_K \to \text{GL}_N(\mathbb{Q}_\ell)$ be an irreducible geometric Galois representation. Then there exists an object $M$ of $\mathcal{M}_{K,\overline{Q}}$ such that

$$r_\ell \otimes_{\mathbb{Q}_\ell} \overline{Q}_\ell \cong H_\iota(M).$$

More generally, the essential image of $H_\iota$ consists of all semisimple geometric representations (with coefficients in $\overline{Q}_\ell$) of $G_K$. 

Proof. The Fontaine–Mazur conjecture asserts that for some smooth projective variety $X/k$, $r_\ell$ is a subquotient of $H^i(X_\mathbb{Q}_\ell, \mathbb{Q}_\ell)(j)$ for some integers $i$ and $j$, and the Grothendieck–Serre conjecture implies this subquotient is in fact a direct summand. Under the Künneth standard conjecture (a consequence of our hypotheses by Lemma 3.2), $\mathcal{M}_K$ has a canonical (weight) grading, and we denote by $H^i(X)$ the weight $i$ component of the motive of $X$. The Tate conjecture then implies (Lemma 3.2) that

$$H_i : \text{End}_{\mathcal{M}_K}(H^i(X)(j)) \otimes_{\mathbb{Q}} \mathbb{Q}_\ell \cong \text{End}_{\mathbb{Q}_\ell[K]}(H^i(X_\mathbb{Q}_\ell, \mathbb{Q}_\ell)(j))$$

(2)
is an isomorphism.

Now, there is a projector (of $\mathbb{Q}_\ell[K]$-modules) $H^i(X_\mathbb{Q}_\ell, \mathbb{Q}_\ell)(j) \rightarrow r_\ell$, which combined with Equation (2) yields a projector in $\text{End}_{\mathcal{M}_K}(H^i(X)(j)) \otimes_{\mathbb{Q}} \mathbb{Q}_\ell$ whose image has $\ell$-adic realization $r_\ell$. But $\text{End}_{\mathcal{M}_K}(H^i(X)(j))$ is a semisimple algebra over $\mathbb{Q}$ (Lemma 3.2), which certainly splits over $\mathbb{Q}$, so the decomposition of $H^i(X)(j)$ into simple objects of $\mathcal{M}_{K,\mathbb{Q}_\ell}$ is already realized in $\mathcal{M}_{K,\overline{\mathbb{Q}}}$.\footnote{In fact, it is realized over the maximal CM subfield of $\overline{\mathbb{Q}}$: see, e.g., [Patrikis 2012, Lemma 4.1.22].}

For the final claim about the essential image (which we do not use in what follows), it suffices to show an irreducible $r_i : G_K \rightarrow \text{GL}_N(\mathbb{Q}_\ell)$ lies in the essential image. Such an $r_i$ is defined over a finite extension of $\mathbb{Q}_\ell$ and can thus be regarded as a higher-dimensional geometric representation $r_\ell$ with $\mathbb{Q}_\ell$-coefficients, necessarily semisimple. By the first part of the lemma, $r_\ell \otimes_{\mathbb{Q}_\ell} \mathbb{Q}_\ell$ is isomorphic to $H_i(M)$ for some $M \in \mathcal{M}_{K,\overline{\mathbb{Q}}}$, and by the Tate conjecture there is a projector in $\text{End}(M) \otimes_{\mathbb{Q}} \mathbb{Q}_\ell$ inducing the canonical (adjunction) projector $r_\ell \otimes_{\mathbb{Q}_\ell} \mathbb{Q}_\ell \twoheadrightarrow r_i$. Arguing as before (a simple object of $\mathcal{M}_{K,\overline{\mathbb{Q}_\ell}}$ arises by scalar-extension from one of $\mathcal{M}_{K,\overline{\mathbb{Q}}}$), we see that $r_i$ is in the essential image of $H_i$. \hfill $\square$

Returning to our particular setting, fix any $\ell_0 \notin S$ as in our first condition on the compatible system $\{\rho_\ell\}_{\ell \notin S}$, and also fix an embedding $i_0 : \overline{\mathbb{Q}} \hookrightarrow \mathbb{Q}_\ell$, so that Lemma 3.3 provides us with a number field (the linear combinations of correspondences needed to cut out a given object of $\mathcal{M}_{K,\overline{\mathbb{Q}}}$ have coefficients in a finite extension of $\mathbb{Q}$) $E \subset \overline{\mathbb{Q}}$ (which we may assume Galois over $\mathbb{Q}$) and a motivic Galois representation $\rho : G_{K,E} \rightarrow \text{GL}_{N,E}$ such that $H_{i_0}(\rho) \cong \rho_{\ell_0} \otimes \mathbb{Q}_\ell$. Let us denote by $\lambda_0$ the place of $E$ induced by $E \subset \overline{\mathbb{Q}} \xrightarrow{i_0} \mathbb{Q}_\ell$. Then for all finite places $\lambda$ of $E$ (say $\lambda | \ell$), and for almost all places $v$ of $K$, compatibility gives us the following equality of rational numbers (note that $\rho_\lambda$ denotes the $\lambda$-adic realization of the motivic Galois representation $\rho$, while $\rho_\ell$ denotes the original $\ell$-adic representation in our compatible system):

$$\text{tr}(\rho_\lambda(\text{fr}_v)) = \text{tr}(\rho_{\lambda_0}(\text{fr}_v)) = \text{tr}(\rho_{\ell_0}(\text{fr}_v)) = \text{tr}(\rho_\ell(\text{fr}_v)).$$
Here we use the fact that the collection of \(\ell\)-adic realizations of a motive form a (weakly) compatible system; this follows from the Lefschetz trace formula, in its “formal” version for correspondences (see for instance [André 2004, 3.3.3, 7.1.4]). We deduce as usual (Brauer–Nesbitt and Chebotarev, see [Serre 1989, theorem on p. I-10; Ribet 1976, Theorem 1.3.1, p. 756]) that \(\rho^{\text{ss}} \otimes_{\mathbb{Q}_\ell} E_{\lambda} \cong \rho_{\lambda}\); this holds for all \(\lambda\) for which \(\rho_\ell\) makes sense, i.e., for all \(\lambda\) above \(\ell \not\in S\).

Recall that for some \(\ell_1 \not\in S\), we have assumed \(\rho_{\ell_1}\) is absolutely irreducible; hence for any place \(\lambda_1\) of \(E\) above \(\ell_1\), the previous paragraph shows that \(\rho_{\lambda_1} \cong \rho_{\ell_1} \otimes E_{\lambda_1}\) is absolutely irreducible. *A fortiori*, \(\rho\) is absolutely irreducible, and then by the Tate conjecture all \(\rho_{\lambda}\) are absolutely irreducible, so we can upgrade the conclusion of the previous paragraph to an isomorphism of absolutely irreducible representations \(\rho_\ell \otimes_{\mathbb{Q}_\ell} E_{\lambda} \cong \rho_{\lambda}\), for all \(\ell \not\in S\).

The next question is whether having each (or almost all) \(\rho_{\lambda}\) in fact definable over \(\mathbb{Q}_\ell\) forces \(\rho\) to be definable over \(\mathbb{Q}\). Since the \(\rho_{\lambda}\) descend to \(\mathbb{Q}_\ell\), the Tate conjecture implies that for all \(\sigma \in \text{Gal}(E/\mathbb{Q})\), \(\sigma \rho \cong \rho\); and since \(\text{End}(\rho)\) is \(E\), the obstruction to descending \(\rho\) to a \(\mathbb{Q}\)-rational representation of \(G_K\) is an element \(\text{obs}_\rho\) of \(H^1(\text{Gal}(E/\mathbb{Q}), \text{PGL}_N(E))\).

**Lemma 3.4.** With the notation above, \(\text{obs}_\rho\) in fact belongs to

\[
\ker \left( H^1(\text{Gal}(E/\mathbb{Q}), \text{PGL}_N(E)) \rightarrow \prod_{\ell \not\in S} H^1(\text{Gal}(E_{\lambda}/\mathbb{Q}_\ell), \text{PGL}_N(E_{\lambda})) \right).
\]

In particular, if \(S\) is empty, then \(\rho\) can be defined over \(\mathbb{Q}\).

**Proof.** We know that each of the \(\lambda\)-adic realizations \(\rho_{\lambda}\) (for \(\lambda | \ell \not\in S\)) can be defined over \(\mathbb{Q}_\ell\); to prove the lemma, we need to verify that the canonical localizations of \(\text{obs}_\rho\) (which arise by extending scalars on the motivic Galois representation) are in fact given by the corresponding obstruction classes for the \(\lambda\)-adic realizations. Thus, we have to recall how these realizations are constructed from \(\rho\) itself. The surjection \(G_K \twoheadrightarrow G_K\) admits a continuous section on \(\mathbb{Q}_\ell\)-points, \(s_\ell : G_K \rightarrow G_K(\mathbb{Q}_\ell)\); composition with \(\rho \otimes_{E} E_{\lambda}\) yields \(\rho_{\lambda}\):

\[
G_K \xrightarrow{s_\ell} G_K(\mathbb{Q}_\ell) \xrightarrow{\rho_{\lambda}} G_K,E(\mathbb{Q}_\ell) \xrightarrow{\rho \otimes_{E} E_{\lambda}} \text{GL}_N(E_{\lambda}).
\]

By construction of the respective obstruction classes, the canonical map from endomorphisms of \(\rho \otimes_{E} E_{\lambda}\) to those of \(\rho_{\lambda}\) realizes the obstruction class for \(\rho_{\lambda}\) as the localization of \(\text{obs}_\rho\) at \(\text{Gal}(E_{\lambda}/\mathbb{Q}_\ell)\). But we have seen that \(\rho_{\lambda}\) can be defined over \(\mathbb{Q}_\ell\), so we conclude that \(\text{obs}_\rho\) has trivial restriction to each \(\text{Gal}(E_{\lambda}/\mathbb{Q}_\ell)\), as desired.
For the final claim, note that by Hilbert 90 we can regard \( \text{obs}_{\rho} \) as an element of
\[
\ker \left( H^2(\text{Gal}(E/\mathbb{Q}), E^\times) \to \prod_{\ell \not\in S} H^2(\text{Gal}(E_\ell/\mathbb{Q}_\ell), E^\times_\ell) \right).
\]
If \( S \) is empty, then the structure of the Brauer group of \( \mathbb{Q} \) (which has only one infinite place!) then forces \( \text{obs}_{\rho} \) to be trivial. \( \square \)

**Proof of Theorem 3.1.** From now on we assume \( S = \emptyset \), so that our compatible system \( \{ \rho_\ell \}_\ell \) arises from a rational representation
\[
\rho : G_K \to \text{GL}_{N, \mathbb{Q}}.
\]
Let \( M \) be the rank \( N \) object of \( \mathcal{M}_K \) corresponding to \( \rho \) via the Tannakian equivalence. Recall that we are given a prime \( \ell_2 \) and a place \( v | \ell_2 \) of \( K \) for which we are given that \( \rho_{\ell_2} | G_v \) is de Rham with Hodge numbers equal to those of an abelian variety of dimension \( N/2 \). All objects of \( \mathcal{M}_K \) enjoy the de Rham comparison theorem of “\( \ell \)-adic Hodge theory”: denoting Fontaine’s period ring over \( K_v \) by \( B_{dR,K_v} \), and the de Rham realization functor by \( H_{dR} : \mathcal{M}_K \to \text{Fil}_K \) (the category of filtered \( K \)-vector spaces), we have the comparison (respecting filtration and \( G_{K_v} \)-action)
\[
H_{dR}(M) \otimes_K B_{dR,K_v} \overset{\sim}{\to} H_{\ell_2}(M) \otimes_{\mathbb{Q}_{\ell_2}} B_{dR,K_v},
\]

hence
\[
H_{dR}(M) \otimes_K K_v \cong D_{dR,K_v}(H_{\ell_2}(M)).
\]
The Hodge filtration on \( H_{dR}(M) \) therefore satisfies
\[
\dim_K \text{gr}^0(H_{dR}(M)) = \dim_K \text{gr}^{-1}(H_{dR}(M)) = \frac{N}{2} \tag{3}
\]
and \( \text{gr}^i(H_{dR}(M)) = 0 \) for \( i \neq 0, -1 \).

Now we turn to the Betti picture. Recall that to define the fiber functor on \( \mathcal{M}_K \) we had to fix an embedding \( K \hookrightarrow \mathbb{C} \); we regard \( K \) as a subfield of \( \mathbb{C} \) via this embedding. Then we also have the analytic Betti–de Rham comparison isomorphism
\[
H_{dR}(M) \otimes_K \mathbb{C} \overset{\sim}{\to} H_B(M|\mathbb{C}) \otimes_{\mathbb{Q}} \mathbb{C}. \tag{4}
\]
We collect our findings in the following lemma, which relies on an application of the Hodge conjecture.

**Lemma 3.5.** There is an abelian variety \( A \) over \( K \), and an isomorphism of motives \( H_1(A) \cong M \).

**Proof.** We see from Equations (3) and (4) that \( H_B(M|\mathbb{C}) \) is a polarizable rational Hodge structure of type \( \{(0, -1), (-1, 0)\} \). It follows from Riemann’s theorem that there is an abelian variety \( A/\mathbb{C} \) and an isomorphism of \( \mathbb{Q} \)-Hodge structures
$H_1(A(\mathbb{C}), \mathbb{Q}) \cong H_B(M|_C)$. The Hodge conjecture implies that this isomorphism comes from an isomorphism $H_1(A) \xrightarrow{\sim} M|_C$ in $\mathcal{M}_C$.

For any $\sigma \in \text{Aut}(\mathbb{C}/\overline{\mathbb{Q}})$, we deduce an isomorphism $^\sigma H_1(A) \xrightarrow{\sim} M|_C \leftarrow H_1(A)$, and again from Riemann’s theorem we see that $^\sigma A$ and $A$ are isogenous.

The following statement will be proven later in this paper.

**Lemma 3.6.** Let $K$ be a countable subfield of the field $\mathbb{C}$ and $\overline{K}$ the algebraic closure of $K$ in $\mathbb{C}$. Let $Y$ be a complex abelian variety of dimension $g$ such that for each field automorphism $\sigma \in \text{Aut}(\mathbb{C}/K)$ the complex abelian variety $Y$ and its “conjugate” $^\sigma Y = Y \times_{\mathbb{C}, \sigma} \mathbb{C}$ are isogenous. Then there exists an abelian variety $Y_0$ over $\overline{K}$ such that $^\sigma Y_0 \times_{\overline{K}} \mathbb{C}$ is isomorphic to $Y$.

It follows from Lemma 3.6 that $A$ has a model $A_{\overline{\mathbb{Q}}}$ over $\overline{\mathbb{Q}}$. The morphism

$$\text{Hom}_{\mathcal{M}_{\overline{\mathbb{Q}}}}(H_1(A_{\overline{\mathbb{Q}}}), M|_{\overline{\mathbb{Q}}}) \rightarrow \text{Hom}_{\mathcal{M}_{\mathbb{C}}}(H_1(A), M|_{\mathbb{C}})$$

is an isomorphism, and then by general principles we deduce the existence of some finite extension $L/K$ inside $\overline{\mathbb{Q}}$ over which $A$ descends to an abelian variety $A_L$, and where we have an isomorphism $H_1(A_L) \xrightarrow{\sim} M|_L$ in $\mathcal{M}_L$.

Finally, we treat the descent to $K$ itself. We form the restriction of scalars abelian variety $\text{Res}_{L/K}(A_L)$; under the fully faithful embedding

$$AV^0_K \subset \mathcal{M}_K, \quad B \mapsto H_1(B),$$

we can think of $H_1(\text{Res}_{L/K}(A_L))$ as $\text{Ind}_L^K(H_1(A_L))$, where the induction is taken in the sense of motivic Galois representations (note that the quotient $G_K/G_L$ is canonically $\text{Gal}(L/K)$, so this is just the usual induction from a finite-index subgroup). Frobenius reciprocity then implies the existence of a nonzero map $M \rightarrow \text{Ind}_L^K(H_1(A_L))$ in $\mathcal{M}_K$. Since $M$ is a simple motive, this map realizes it as a direct summand in $\mathcal{M}_K$, and consequently (full-faithfulness) in $AV^0_K$ as well. That is, there is an endomorphism of $\text{Res}_{L/K}(A_L)$ whose image is an abelian variety $A$ over $K$ with $H_1(A) \cong M$.

**Proof of Lemma 3.6.** We may assume that $g \geq 1$. Since $\overline{K}$ is also countable, we may replace $K$ by $\overline{K}$, i.e., assume that $K$ is algebraically closed. Since the isogeny class of $Y$ consists of a countable set of (complex) abelian varieties (up to an isomorphism), we conclude that the set $\text{Aut}(\mathbb{C}/K)(Y)$ of isomorphism classes of complex abelian varieties of the form $[^\sigma Y | \sigma \in \text{Aut}(\mathbb{C}/K)]$ is either finite or countable.

Our plan is as follows. Let us consider a fine moduli space $A_{g,?}$ over $\overline{\mathbb{Q}}$ of $g$-dimensional abelian varieties (schemes) with certain additional structures (there should be only finitely many choices of these structures for any given abelian variety) such that it is a quasiprojective subvariety in some projective space $\mathbb{P}^N$.

...
Choose these additional structures for \( \mathcal{Y} \) (there should be only finitely many choices) and let \( P \in \mathcal{A}_g,?^{\mathcal{Y}}(\mathbb{C}) \) be the corresponding point of our moduli space. We need to prove that

\[
P \in \mathcal{A}_g,?^{\mathcal{Y}}(K).
\]

Suppose that it is not true. Then the orbit \( \text{Aut}(\mathbb{C}/K)(P) \) of \( P \) is uncountable. Indeed, \( P \) lies in one of the \((N+1)\) affine charts/spaces \( \mathbb{A}^N \) that do cover \( \mathbb{D}^N \). This implies that \( P \) does not belong to \( \mathbb{A}^N(K) \) and therefore (at least) one of its coordinates is transcendental over \( K \), and \( \text{Aut}(\mathbb{C}/K)(P) \) coincides with uncountable \( \mathbb{C} \setminus \overline{K} \) and therefore the \( \text{Aut}(\mathbb{C}/K)(P) \) of \( P \) is uncountable in \( \mathcal{A}_g,?^{\mathcal{Y}}(\mathbb{C}) \). However, for each \( \sigma \in \text{Aut}(\mathbb{C}/K) \) the point \( \sigma(P) \) corresponds to \( \sigma\mathcal{Y} \) with some additional structures and there are only finitely many choices for these structures. Since we know that the orbit \( \text{Aut}(\mathbb{C}/K)(\mathcal{Y}) \) of \( \mathcal{Y} \), is, at most, countable, we conclude that the orbit \( \text{Aut}(\mathbb{C}/K)(P) \) of \( P \) is also, at most, countable, which is not the case. This gives us a desired contradiction.

We choose as \( \mathcal{A}_g,? \) the moduli space of (polarized) abelian schemes of relative dimension \( g \) with theta structures of type \( \delta \) that was introduced and studied by D. Mumford [1966]. In order to choose (define) a suitable \( \delta \), let us pick a totally symmetric ample invertible sheaf \( L_0 \) on \( \mathcal{Y} \) [Mumford 1966, Section 2] and consider its 8th power \( L := L_0^8 \) in \( \text{Pic}(\mathcal{Y}) \). Then \( L \) is a very ample invertible sheaf that defines a polarization \( \Lambda(L) \) on \( \mathcal{Y} \) [Mumford 1966, Part I, Section 1] that is an isogeny from \( \mathcal{Y} \) to its dual; the kernel \( H(L) \) of \( \Lambda(L) \) is a finite commutative subgroup of \( \mathcal{Y}(\mathbb{C}) \) (that contains all points of order 8). The order of \( H(L) \) is the degree of the polarization. The type \( \delta \) is essentially the isomorphism class of the group \( H(L) \) [Mumford 1966, Part I, Section 1, p. 294]. The resulting moduli space \( \mathcal{A}_g,?^{\mathcal{Y}} := M_\delta \) [Mumford 1966, Part II, Section 6] enjoys all the properties that we used in the course of the proof.

Here is the anabelian application already mentioned in the introduction:

**Theorem 3.7.** Assume the Hodge, Tate, Fontaine–Mazur, and Grothendieck–Serre conjectures. Suppose \( s \in S_0(K, \mathcal{A}_g) \) gives rise to a system of \( \ell \)-adic Galois representations one of which is absolutely irreducible. Then there exists up to isomorphism a unique principally polarized abelian variety \( B/K \) with \( \sigma_{\mathcal{A}_g/K}(B) = s \).

**Proof.** Let us write \( s_\ell \) for the \( \ell \)-adic representation associated to \( s \); thus \( s_\ell \) is a representation of \( G_K \) on a free \( \mathbb{Z}_\ell \)-module \( T_\ell \) of rank \( 2g \), automatically satisfying Hypothesis 2 of Theorem 3.1 since \( s \) belongs to \( S_0(K, \mathcal{A}_g) \). Hypothesis 1 of Theorem 3.1 is satisfied by assumption, so we obtain an abelian variety \( A/K \) (well-defined up to isogeny) whose rational Tate modules \( V_\ell(A) \) are isomorphic (as \( \ell \)-adic representations) to the given \( s_\ell \otimes_{\mathbb{Z}_\ell} \mathbb{Q}_\ell \) (for all \( \ell \)). Moreover Hypothesis 1 implies that the endomorphism ring of \( A \) is \( \mathbb{Z} \). It remains to see that within the isogeny class
of $A$ there is a \textit{principally polarized} abelian variety $B$ over $K$ whose integral Tate module $T_\ell(B)$ is isomorphic as a $\mathbb{Z}_\ell[G_K]$-module to $T_\ell$ (for all $\ell$), i.e., such that $\sigma_{A_\ell/K}(B) = s$. For this, we first observe that by [Deligne 1971, Proposition 3.3] (which readily generalizes to abelian varieties of any dimension), it suffices to show that for almost all $\ell$, there is an isomorphism $T_\ell(A) \cong T_\ell$. Since $\text{End}(A) = \mathbb{Z}$, [Zarhin 1985, Corollary 5.4.5] implies that $A[\ell]$ is an absolutely simple Galois module for almost all $\ell$, and hence that for almost all $\ell$, all Galois-stable lattices in $V_\ell(A)$ are of the form $\ell^m T_\ell(A)$ for some integer $m$; we conclude that $T_\ell(A)$ is isomorphic to $T_\ell$ for almost all $\ell$. Thus there exists an abelian variety $B$ in the isogeny class of $A$ such that the $\mathbb{Z}_\ell[G_K]$-modules $T_\ell(B)$ and $T_\ell$ are isomorphic for all $\ell$.

In order to prove the uniqueness of such a $B$ up to an isomorphism, first, notice that $\text{End}(B) = \mathbb{Z}$. Second, let $C$ be an abelian variety over $K$ such that the $\mathbb{Z}_\ell[G_K]$-modules $T_\ell(B)$ and $T_\ell(C)$ are isomorphic for all primes $\ell$. This implies that the $\mathbb{Z}_\ell$-ranks of $T_\ell(B)$ and $T_\ell(C)$ coincide and therefore

$$\dim(B) = \dim(C).$$

By a theorem of Faltings [1983],

$$\text{Hom}(B, C) = \text{Hom}_{G_\ell}(T_\ell(B), T_\ell(C)).$$

Since $\text{Hom}(B, C)$ is dense in $\text{Hom}(B, C) \otimes \mathbb{Z}_\ell$ in the $\ell$-adic topology, and the set of isomorphisms $T_\ell(B) \cong T_\ell(C)$ is open in $\text{Hom}(B, C) \otimes \mathbb{Z}_\ell$, there is a homomorphism $\phi_\ell \in \text{Hom}(B, C)$ that induces an isomorphism of Tate modules $T_\ell(B) \cong T_\ell(C)$. Clearly, ker($\phi_\ell$) does not contain points of order $\ell$ and therefore is finite. Since $\dim(B) = \dim(C)$, we obtain that $\phi_\ell$ is an isogeny, whose degree is prime to $\ell$. In particular, $B$ and $C$ are isogenous. On the other hand, since $\text{End}(B) = \mathbb{Z}$, the group $\text{Hom}(B, C)$ is a free $\mathbb{Z}$-module of rank 1. Let us choose $\psi : B \rightarrow C$ that is a generator of $\text{Hom}(B, C)$. Clearly, $\psi$ is an isogeny. Since for \textit{all primes} $\ell$

$$\phi_\ell \in \text{Hom}(B, C) = \mathbb{Z} \cdot \psi,$$

deg($\psi$) is not divisible by $\ell$ and therefore deg($\psi$) = 1, i.e., $\psi$ is an isomorphism of abelian varieties $B$ and $C$.

We still need to check that $B$ is \textit{principally polarized}. Since $s_\ell$ comes from $s$, there is an \textit{alternating} Galois-equivariant $\mathbb{Z}_\ell$-bilinear perfect/unimodular form

$$T_\ell \times T_\ell \rightarrow \mathbb{Z}_\ell(1).$$

Since $T_\ell$ is isomorphic as a $\mathbb{Z}_\ell[G_K]$-module to $T_\ell(B)$, there is a Galois-equivariant, $\mathbb{Z}_\ell$-bilinear perfect/unimodular form

$$T_\ell(B) \times T_\ell(B) \rightarrow \mathbb{Z}_\ell(1).$$
This implies that the Galois modules $T_\ell(B)$ and $\text{Hom}_{\mathbb{Z}_\ell}(T_\ell(B), \mathbb{Z}_\ell(1))$ are isomorphic. It follows from the last sentence of Section 2 that the Galois modules $T_\ell(B)$ and $T_\ell(\hat{B})$ are isomorphic for all primes $\ell$. This implies that the abelian varieties $\hat{B}$ and $B$ are isomorphic. Since $\text{End}(B) = \mathbb{Z}$, there is an isomorphism $\mu : B \to \hat{B}$ such that $\text{Hom}(B, \hat{B}) = \mathbb{Z} \cdot \mu$. Let $\lambda : B \to \hat{B}$ be a polarization on $B$. Then there is a nonzero integer $n$ such that $\lambda = n \cdot \mu$. Replacing if necessary $\mu$ by $-\mu$, we may and will assume that $n$ is a positive integer. It follows from [Mumford 1970, Section 23, Theorem 3] that $\mu$ is a polarization, which is obviously principal. (Clearly, there is exactly one principal polarization on $B$, namely $\mu$.) So, $\sigma_{A_s/K}(B)$ is defined and obviously coincides with $s$. □

Remark 3.8. Note that for each prime $\ell$ we get the Riemann form [Lang 1959, Chapter VII, Section 2; Mumford 1970, Section 20]

$$E_{\ell,\mu} : T_\ell(B) \times T_\ell(B) \to \mathbb{Z}_\ell(1), \quad x, y \mapsto e_\ell(x, \mu y) \quad \text{for all} \quad x, y \in T_\ell(B),$$

which is an alternating Galois-equivariant $\mathbb{Z}_\ell$-bilinear perfect/unimodular form on the free $\mathbb{Z}_\ell$-module $T_\ell(B)$. Since $\text{End}(B) = \mathbb{Z}$, the already cited result of Faltings implies that $\text{End}_G_{K}(T_\ell(A)) = \mathbb{Z}_\ell$. It follows that any alternating Galois-equivariant $\mathbb{Z}_\ell$-bilinear perfect/unimodular form

$$T_\ell(B) \times T_\ell(B) \to \mathbb{Z}_\ell(1)$$

coincides with $c_\ell \cdot E_{\ell,\mu}$ for some $c_\ell \in \mathbb{Z}_\ell^*$. This implies that any isomorphism between the $\mathbb{Z}_\ell[G_{K}]$-modules $T_\ell$ and $T_\ell(B)$ induces isomorphisms between the corresponding symplectic groups and between the corresponding groups of symplectic similitudes.

Results in the same vein as this corollary have been obtained for elliptic curves over $\mathbb{Q}$ in [Helm and Voloch 2011] and for elliptic curves over function fields in [Voloch 2012].

4. Counterexamples

Now we will construct an example of Galois representation that will provide us with examples that show that some of the hypotheses of the above results are indispensable.

Let $k$ be a real quadratic field. Let us choose a prime $p$ that splits in $k$. Now let $D$ be the indefinite quaternion $k$-algebra that splits everywhere outside (two) prime divisors of $p$ and is ramified at these divisors. If $\ell$ is a prime then we have

$$D \otimes_{\mathbb{Q}} \mathbb{Q}_\ell = [D \otimes_k k] \otimes_{\mathbb{Q}} \mathbb{Q}_\ell = D \otimes_k [k \otimes_{\mathbb{Q}} \mathbb{Q}_\ell].$$

This implies that if $\ell \neq p$ then $D \otimes_{\mathbb{Q}} \mathbb{Q}_\ell$ is either (isomorphic to) the simple matrix algebra (of size 2) over a quadratic extension of $\mathbb{Q}_\ell$ or a direct sum of two copies of
the simple matrix algebra (of size $2$) over $\mathbb{Q}_\ell$. (In both cases, $D \otimes_{\mathbb{Q}} \mathbb{Q}_\ell$ is isomorphic to the matrix algebra $M_2(k \otimes_{\mathbb{Q}} \mathbb{Q}_\ell)$ of size $2$ over $k \otimes_{\mathbb{Q}} \mathbb{Q}_\ell$.)

In particular, the image of $D \otimes_{\mathbb{Q}} \mathbb{Q}_\ell$ under each nonzero $\mathbb{Q}_\ell$-algebra homomorphism contains zero divisors.

Let $Y$ be an abelian variety over a field $L$. Suppose that all $\bar{L}$-endomorphisms of $Y$ are defined over $L$ and there is a $\mathbb{Q}$-algebra embedding

$$D \hookrightarrow \text{End}^0(Y)$$

that sends $1$ to $1$. This gives us the embedding

$$D \otimes_{\mathbb{Q}} \mathbb{Q}_\ell \subset \text{End}^0(Y) \otimes_{\mathbb{Q}} \mathbb{Q}_\ell \subset \text{End}_{G_\ell}(V_\ell(Y)).$$

Recall that if $\ell \neq p$ then $D \otimes_{\mathbb{Q}} \mathbb{Q}_\ell$ is isomorphic to the matrix algebra of size $2$ over $k \otimes_{\mathbb{Q}} \mathbb{Q}_\ell$. This implies that there are two isomorphic $\mathbb{Q}_\ell[G_L]$-submodules $W_{1,\ell}(Y)$ and $W_{2,\ell}(Y)$ in $V_\ell(Y)$ such that

$$V_\ell(Y) = W_{1,\ell}(Y) \oplus W_{2,\ell}(Y) \cong W_{1,\ell}(Y) \oplus W_{2,\ell}(Y) \cong W_{2,\ell}(Y) \oplus W_{2,\ell}(Y).$$

If we denote by $W_\ell(Y)$ the $\mathbb{Q}_\ell[G_L]$-module $W_{1,\ell}$ then we get an isomorphism of $\mathbb{Q}_\ell[G_L]$-modules

$$V_\ell(Y) \cong W_\ell(Y) \oplus W_\ell(Y).$$

This implies that the centralizer $\text{End}_{G_\ell}(V_\ell(Y))$ coincides with the matrix algebra $M_2(\text{End}_{G_\ell}(W_\ell(Y)))$ of size $2$ over the centralizer $\text{End}_{G_\ell}(W_\ell(Y))$.

If $\ell = p$ then $k \otimes_{\mathbb{Q}} \mathbb{Q}_p = \mathbb{Q}_p \oplus \mathbb{Q}_p$ and $D \otimes_{\mathbb{Q}} \mathbb{Q}_p$ splits into a direct sum of two (mutually isomorphic) quaternion algebras over $\mathbb{Q}_p$. This also gives us a splitting of the Galois module $V_p(Y)$ into a direct sum

$$V_p(Y) = W_{1,p}(Y) \oplus W_{2,p}(Y).$$

of its certain nonzero $\mathbb{Q}_p[G_L]$-submodules $W_{1,p}(Y)$ and $W_{2,p}(Y)$. (Actually,

$$\dim_{\mathbb{Q}_p} W_{1,p} = \dim_{\mathbb{Q}_p} W_{2,p} = \dim(Y),$$

because $V_p(Y)$ is a free $k \otimes_{\mathbb{Q}} \mathbb{Q}_p$-module of rank $2 \dim(Y)/[k : \mathbb{Q}] = \dim(Y)$ [Ribet 1976, Theorem 2.1.1 on p. 768].)

**Remark.** Let $L$ be a finitely generated field of characteristic $0$. Suppose that $D = \text{End}^0(Y)$. By Faltings’ results [1983; 1984] about the Galois action on Tate modules of abelian varieties, the $G_L$-module $V_\ell(Y)$ is semisimple and

$$\text{End}_{G_\ell}(V_\ell(Y)) = D \otimes_{\mathbb{Q}} \mathbb{Q}_\ell.$$ 

This implies that if $\ell \neq p$ then (the submodule) $W_\ell(Y)$ is also semisimple and

$$M_2(\text{End}_{G_\ell}(W_\ell(Y))) \cong M_2(k \otimes_{\mathbb{Q}} \mathbb{Q}_\ell).$$
It follows that
\[ \text{End}_{G_L}(W_\ell(Y)) \cong k \otimes \mathbb{Q} \ell. \]

On the other hand, the $G_L$-modules $W_{1,p}(Y)$ and $W_{2,p}(Y)$ are nonisomorphic.

According to Shimura [1963] (see also the case of Type II($e_0 = 2$) with $m = 1$ in [Oort 1988, Table 8.1 on p. 498] and [Oort and Zarhin 1995, table on p. 23]), there exists a complex abelian fourfold $X$, whose endomorphism algebra $\text{End}^0(X)$ is isomorphic to $D$. Clearly, $X$ is defined over a finitely generated field of characteristic zero. It follows from Serre’s variant of Hilbert’s irreducibility theorem for infinite Galois extensions combined with results of Faltings that there exists a number field $K$ and an abelian fourfold $A$ over $K$ such that the endomorphism algebra of all $\overline{K}$-endomorphisms of $A$ is also isomorphic to $D$ (see [Noot 1995, Corollary 1.5 on p. 165]). Enlarging $K$, we may assume that all points of order 12 on $A$ are defined over $K$ and therefore, by the theorem of Silverberg, all $\overline{K}$-endomorphisms of $A$ are defined over $K$. Now Raynaud’s criterion [SGA 71 1972] (see also [Silverberg and Zarhin 1995]), implies that $A$ has everywhere semistable reduction. On the other hand,

\[ \dim \mathbb{Q} \text{End}^0(A) = 8 > 4 = \dim(A). \]

By [Oort 1988, Lemma 3.9 on p. 484], $A$ has everywhere potential good reduction. This implies that $A$ has good reduction everywhere. If $v$ is a nonarchimedean place of $K$ with finite residue field $\kappa(v)$ then we write $A(v)$ for the reduction of $A$ at $v$; clearly, $A(v)$ is an abelian fourfold over $\kappa(v)$. If $\text{char}(\kappa(v)) \neq 2$ then all points of order 4 on $A(v)$ are defined over $\kappa(v)$; if $\text{char}(\kappa(v)) \neq 3$ then all points of order 3 on $A(v)$ are defined over $\kappa(v)$. It follows from the theorem of Silverberg that all $\kappa(v)$-endomorphisms of $A(v)$ are defined over $\kappa(v)$. For each $v$ we get an embedding of $\mathbb{Q}$-algebras

\[ D \cong \text{End}^0(A) \hookrightarrow \text{End}^0(A(v)). \]

In particular, $\text{End}^0(A(v))$ is a noncommutative $\mathbb{Q}$-algebra, whose $\mathbb{Q}$-dimension is divisible by 8.

**Theorem 4.1.** If $\ell := \text{char}(\kappa(v)) \neq p$ then $A(v)$ is not simple over $\kappa(v)$.

**Proof.** We write $q_v$ for the cardinality of $\kappa(v)$. Clearly, $q_v$ is a power of $\ell$.

Suppose that $A(v)$ is simple over $\kappa(v)$. Since all endomorphisms of $A(v)$ are defined over $\kappa(v)$, the abelian variety $A(v)$ is absolutely simple.

Let $\pi$ be a Weil $q_v$-number that corresponds to the $\kappa(v)$-isogeny class of $A(v)$ [Tate 1966; 1971]. In particular, $\pi$ is an algebraic integer (complex number), all whose Galois conjugates have (complex) absolute value $\sqrt{q_v}$. In particular, the product

\[ \pi \overline{\pi} = q_v, \]
Anabelian geometry and descent obstructions on moduli spaces

where $\pi$ is the complex conjugate of $\pi$.

Let $E = \mathbb{Q}(\pi)$ be the number field generated by $\pi$ and let $\mathcal{O}_E$ be the ring of integers in $E$. Then $E$ contains $\pi$ and is isomorphic to the center of $\text{End}^0(A(v))$ [Tate 1966; 1971]; one may view $\text{End}^0(A(v))$ as a central division algebra over $E$. It is known that $E$ is either $\mathbb{Q}$, $\mathbb{Q}(\sqrt{\ell})$ or a (purely imaginary) CM field [Tate 1971, p. 97]. It is known [ibid] that in the first two (totally real) cases simple $A(v)$ has dimension 1 or 2, which is not the case. So, $E$ is a CM field; Since $\dim(A(v)) = 4$ and $[E : \mathbb{Q}]$ divides $2 \dim(A(v))$, we have $[E : \mathbb{Q}] = 2$, 4 or 8. By [Tate 1971, p. 96, Theorem 1(ii), formula (2)]\(^3\),

$$8 = 2 \cdot 4 = 2 \dim(A(v))) = \sqrt{\dim_E(\text{End}^0(A(v))) \cdot [E : \mathbb{Q}]}.$$

Since $\text{End}^0(A(v))$ is noncommutative, it follows that $E$ is either an imaginary quadratic field and $\text{End}^0(A(v))$ is a 16-dimensional division algebra over $E$ or $E$ is a CM field of degree 4 and $\text{End}^0(A(v))$ is a 4-dimensional (i.e., quaternion) division algebra over $E$. In both cases $\text{End}^0(A(v))$ is unramified at all places of $E$ except some places of residual characteristic $\ell$ [Tate 1971, p. 96, Theorem 1(ii)]. It follows from the Hasse–Brauer–Noether theorem that $\text{End}^0(A(v))$ is ramified at, at least, two places of $E$ with residual characteristic $\ell$. This implies that $\mathcal{O}_E$ contains, at least, two maximal ideals that lie above $\ell$.

Clearly,

$$\pi, \pi \in \mathcal{O}_E.$$

Recall that $\pi \pi = q_v$ is a power of $\ell$. This implies that for every prime $r \neq \ell$ both $\pi$ and $\pi$ are $r$-adic units in $E$.

First assume that $E$ has degree 4 and $\text{End}^0(A(v))$ is a quaternion algebra. Then (thanks to the theorem of Hasse–Brauer–Noether) there exists a place $w$ of $E$ with residual characteristic $\ell$ and such that the localization $\text{End}^0(A(v)) \otimes_E E_w$ is a quaternion division algebra over the $w$-adic field $E_w$. On the other hand, there is a nonzero (because it sends 1 to 1) $\mathbb{Q}_\ell$-algebra homomorphism

$$D \otimes_{\mathbb{Q}} \mathbb{Q}_\ell \to \text{End}^0(A(v)) \otimes_{\mathbb{Q}} \mathbb{Q}_\ell \to \text{End}^0(A(v)) \otimes_E E_w.$$

This implies that $\text{End}^0(A(v)) \otimes_E E_w$ contains zero divisors, which is not the case and we get a contradiction.

So, now we assume that $E$ is an imaginary quadratic field and

$$\dim_E(\text{End}^0(A(v))) = 16 = 4^2.$$

In particular, the order of the class of $\text{End}^0(A(v))$ in the Brauer group of $E$ divides 4 and therefore is either 2 or 4.

\(^3\)In [Tate 1971] our $E$ is denoted by $F$ while our $\text{End}^0(A(v))$ is denoted by $E$. 
We have already seen that there exist, at least, two maximal ideals in \( \mathcal{O}_E \) that lie above \( \ell \). Since \( E \) is an imaginary quadratic field, the ideal \( \ell \mathcal{O}_L \) of \( \mathcal{O}_L \) splits into a product of two distinct complex-conjugate maximal ideals \( w_1 \) and \( w_2 \) and therefore

\[
E_{w_1} = \mathbb{Q}_\ell, \quad E_{w_2} = \mathbb{Q}_\ell; \quad [E_{w_1} : \mathbb{Q}_\ell] = [E_{w_2} : \mathbb{Q}_\ell] = 1.
\]

Let

\[
\text{ord}_{w_i} : E^* \twoheadrightarrow \mathbb{Z}
\]

be the discrete valuation map that corresponds to \( w_i \). Recall that \( q_v \) is a power of \( \ell \), i.e., \( q_v = \ell^N \) for a certain positive integer \( N \). Clearly

\[
\text{ord}_{w_i}(\ell) = 1, \quad \text{ord}_{w_i}(\pi) + \text{ord}_{w_i}(\overline{\pi}) = \text{ord}_{w_i}(q_v) = N.
\]

By [Tate 1971, p. 96, Theorem 1(ii), formula (1)], the local invariant of \( \text{End}^0(A(v)) \) at \( w_i \) is

\[
\frac{\text{ord}_{w_i}(\pi)}{\text{ord}_{w_i}(q_v)} \cdot [E_{w_i} : \mathbb{Q}_\ell] \pmod{1} = \frac{\text{ord}_{w_i}(\pi)}{N} \pmod{1}.
\]

In addition, the sum in \( \mathbb{Q}/\mathbb{Z} \) of local invariants of \( \text{End}^0(A(v)) \) at \( w_1 \) and \( w_2 \) is zero [Tate 1971, Section 1, Theorem 1 and Example b]; we have already seen that its local invariants at all other places of \( E \) do vanish. Using the Hasse–Brauer–Noether theorem and taking into account that the order of the class of \( \text{End}^0(A(v)) \) in the Brauer group of \( E \) is either 2 or 4, we conclude that the local invariants of \( \text{End}^0(A(v)) \) at \( \{w_1, w_2\} \) are either \( \{\frac{1}{4} \pmod{1}, \frac{3}{4} \pmod{1}\} \) or \( \{\frac{3}{4} \pmod{1}, \frac{1}{4} \pmod{1}\} \) (and in both cases the order of \( \text{End}^0(A(v)) \) in the Brauer group of \( E \) is 4) or \( \{\frac{1}{2} \pmod{1}, \frac{1}{2} \pmod{1}\} \). In the latter case it follows from the formula for the \( w_i \)-adic invariant of \( \text{End}^0(A(v)) \) that

\[
\text{ord}_{w_i}(\pi) = \frac{N}{2} = \text{ord}_{w_i}(\overline{\pi})
\]

and therefore \( \overline{\pi}/\pi \) is a \( w_i \)-adic unit for both \( w_1 \) and \( w_2 \). Therefore \( \overline{\pi}/\pi \) is an \( \ell \)-adic unit. This implies that \( \overline{\pi}/\pi \) is a unit in imaginary quadratic \( E \) and therefore is a root of unity. It follows that

\[
\frac{\pi^2}{q_v} = \frac{\pi^2}{\pi \overline{\pi}} = \frac{\pi}{\overline{\pi}}
\]

is a root of unity. This implies that there is a positive (even) integer \( m \) such that

\[
\pi^m = q_v^{m/2} \in \mathbb{Q}
\]

and therefore \( \mathbb{Q}(\pi^m) = \mathbb{Q} \). Let \( \kappa(v)_m \) be the finite degree \( m \) field extension of \( \kappa(v) \), which consists of \( q_v^m \) elements. Then \( \pi^m \) is the Weil \( q_v^m \)-number that corresponds to the simple 4-dimensional abelian variety \( A(v) \times \kappa(v)_m \) over \( \kappa(v)_m \). Since \( \mathbb{Q}(\pi^m) = \mathbb{Q} \), we conclude (as above) that \( A(v) \times \kappa(v)_m \) has dimension 1 or 2, which is not the case.
In both remaining cases the order of the algebra $\text{End}^0(A(v)) \otimes_E E_{w_1}$ in the Brauer group of the $E_{w_1} \cong \mathbb{Q}_\ell$ is 4. This implies that $\text{End}^0(A(v)) \otimes_E E_{w_1}$ is neither the matrix algebra of size 4 over $E_{w_1}$ nor the matrix algebra of size two over a quaternion algebra over $E_{w_1}$. The only remaining possibility is that $\text{End}^0(A(v)) \otimes_E E_{w_1}$ is a division algebra over $E_{w_1}$. However, there is again a nonzero (because it sends 1 to 1) $\mathbb{Q}_\ell$-algebra homomorphism

$$D \otimes \mathbb{Q}_\ell \rightarrow \text{End}^0(A(v)) \otimes \mathbb{Q}_\ell \rightarrow \text{End}^0(A(v)) \otimes E_{w_1}.$$ 

This implies that $\text{End}^0(A(v)) \otimes_E E_{w_1}$ contains zero divisors, which is not the case and we get a contradiction. \hfill \Box

**Theorem 4.2.** If $\ell := \text{char}(\kappa(v)) \neq p$ then there exists an abelian surface $B(v)$ over $\kappa(v)$ such that $A(v)$ is $\kappa(v)$-isogenous to the square $B(v)^2$ of $B(v)$.

**Proof.** We know that $A(v)$ is not simple and that all $\overline{\kappa(v)}$-endomorphisms of $A(v)$ are defined over $k(v)$. Now let us split $A(v)$ up to a $\kappa(v)$-isogeny into a product of its $\kappa(v)$-isotypic components, using the Poincaré complete reducibility theorem [Lang 1959, Theorem 6 on p. 28 and Theorem 7 on p. 30]. In other words, there is a $\kappa(v)$-isogeny

$$S : \prod_{i \in I} A_i \rightarrow A(v),$$

where each $A_i$ is a nonzero abelian $\kappa(v)$-subvariety in $A$ such that $\text{End}^0(A_i)$ is a simple $\mathbb{Q}$-algebra and $S$ induces an isomorphism of $\mathbb{Q}$-algebras

$$\text{End}^0(A(v)) \cong \text{End}^0\left(\prod_{i \in I} A_i\right) = \bigoplus_{i \in I} \text{End}^0(A_i).$$

This gives us nonzero $\mathbb{Q}$-algebra homomorphisms

$$D \rightarrow \text{End}^0(A_i)$$

that must be injective, since $D$ is a simple $\mathbb{Q}$-algebra. This implies that each $\text{End}^0(A_i)$ is a noncommutative simple $\mathbb{Q}$-algebra, whose $\mathbb{Q}$-dimension is divisible by 8. In particular, all $\dim(A_i) \geq 2$ and therefore $I$ consists of, at most, 2 elements, since

$$\sum_{i \in I} \dim(A_i) = \dim(A(v)) = 4.$$

Since all $\overline{\kappa(v)}$-endomorphisms of $A(v)$ are defined over $k(v)$, all $\overline{\kappa(v)}$-endomorphisms of $A_i$ are also defined over $\kappa(v)$; in addition, if $i$ and $j$ are distinct elements of $I$, then every $\overline{\kappa(v)}$-homomorphism between $A_i$ and $A_j$ is 0.

If we have $\dim(A_i) = 2$ for some $i$ then either $A_i$ is isogenous to a square of a supersingular elliptic curve or $A_i$ is an absolutely simple abelian surface. However,
each absolutely simple abelian surface over a finite field is either ordinary (i.e., the slopes of its Newton polygon are 0 and 1, both of length 2) or almost ordinary (i.e., the slopes of its Newton polygon are 0 and 1, both of length 1, and $\frac{1}{2}$ with length 2): this assertion is well known and follows easily from [Zarhin 2015, Remark 4.1 on p. 2088]. However, in both (ordinary and almost ordinary) cases the endomorphism algebra of a simple abelian variety is commutative [Oort 1992, Lemma 2.3 on p. 136]. This implies that if $\dim(A_i) = 2$ then $A_i$ is $\kappa(v)$-isogenous to a square of a supersingular elliptic curve. However, if $I$ consists of two elements, say $i$ and $j$, then it follows that both $A_i$ and $A_j$ are 2-dimensional and therefore both isogenous to a square of a supersingular elliptic curve. This implies that $A_i$ and $A_j$ are isotypic and therefore $A$ itself is isotypic and we get a contradiction, i.e., none of the $A_i$ has dimension 2. It is also clear that if $\dim(A_j) = 3$ then $\dim(A_j) = 1$, which could not be the case. This implies that $A(v)$ itself is isotypic. It follows that if $\ell = \text{char}(\kappa(v)) \neq p$ then $A(v)$ is $\kappa(v)$-isogenous either to a 4th power of an elliptic curve or to a square of an abelian surface over $\kappa(v)$. (Recall that $A(v)$ is not simple!) In both cases there exists an abelian surface $B(v)$ over $\kappa(v)$, whose square $B(v)^2$ is $\kappa(v)$-isogenous to $A(v)$.

Let $B(v)$ be as in Theorem 4.2. One may lift the abelian surface $B(v)$ over $\kappa(v)$ to an abelian surface $B^v$ over $K_v$, whose reduction is $B(v)$ (see [Oort 1987, Proposition 11.1 on p. 177]). Now if one restricts the action of $G_K$ on the $\mathbb{Q}_r$-Tate module (here $r$ is any prime different from $\text{char}(\kappa(v))$)

$$V_r(A) = T_r(A) \otimes_{\mathbb{Z}_r} \mathbb{Q}_r$$

to the decomposition group $D(v) = G_{K_v}$, then the corresponding $G_{K_v}$-module $V_r(A)$ is unramified (i.e., the inertia group acts trivially) and isomorphic to

$$V_r(B^v) \oplus V_r(B^v).$$

**Theorem 4.3.** If $r \neq p$ and $\text{char}(\kappa(v)) \neq r$ then the $G_{K_v}$-modules $V_r(B^v)$ and $W_r(A)$ are isomorphic. In particular, the $G_{K_v}$-modules

$$V_r(A) = W_r(A) \oplus W_r(A)$$

and

$$V_r(B^v) \oplus V_r(B^v) = V_r((B^v)^2)$$

are isomorphic.

**Proof.** We know that the $G_{K_v}$-modules $W_r(A) \oplus W_r(A)$ and

$$V_r(B^v) \oplus V_r(B^v)$$

are both isomorphic to $V_r(A)$. Since the frobenius endomorphism of $A(v)$ acts on $V_r(A)$ as a semisimple linear operator (by a theorem of A. Weil), the $G_{K_v}$-module
$V_r(A)$ is semisimple. This implies that the $G_{K_v}$-modules $V_r(B^v)$ and $W_r(A)$ are isomorphic. $\square$

For primes $\ell \neq p$, the algebra $D \otimes_{Q} Q_\ell$ splits, and correspondingly, the representation $V_\ell(A)$ splits as $W_\ell \oplus W_\ell$. Locally, at a place $v \nmid \ell$, we have $W_\ell \cong V_\ell(B^v)$. However, globally, the representation $W_\ell$ does not come from an abelian variety over $K$. Indeed, if the $G_K$-module $W_\ell$ is isomorphic to $V_\ell(B)$ for an abelian variety $B$ over $K$ then $\dim(B) = 2$ and the theorem of Faltings implies that there is a nonzero homomorphism of abelian varieties $B \to A$ over $K$, which is not the case, since the fourfold $A$ is simple. On the other hand, if $v|\ell$ then $V_\ell(A)$ is a de Rham representation of $G_{K_v}$ with weights 0 and $-1$, both of multiplicity $\dim(A) = 4$. Since a subrepresentation of a de Rham representation is also de Rham, we conclude that $W_\ell$ is de Rham. It is also clear that $W_\ell$ has the same Hodge–Tate weights as

$$V_\ell(A) = W_\ell \oplus W_\ell$$

but the multiplicities should be divided by 2, i.e., the Hodge–Tate weights of $W_\ell$ are 0 and $-1$, both of multiplicity 2.

We thus obtain:

**Theorem 4.4.** The system of representations $\{W_\ell\}_{\ell \neq p}$ constructed above does not come globally from an abelian variety defined over the field $K$ but for all $v \nmid \ell$ the representation $W_\ell$ locally comes from an abelian variety $B^v/K_v$. In particular, $\{W_\ell\}_{\ell \neq p}$ is a weakly compatible system of 4-dimensional $\ell$-adic representations of $G_K$.

If $v|\ell$ then $W_\ell$ is locally a de Rham representation with Hodge–Tate weights 0 and $-1$, both of multiplicity 2.

**Remark.** By a theorem of Faltings [1983], the $G_K$-module $V_\ell(A)$ is semisimple and therefore its submodule $W_\ell$ is also semisimple. On the other hand, we know that the centralizer

$$\text{End}_{G_K}(W_\ell) = k \otimes_{Q} Q_\ell \neq Q_\ell;$$

in particular, none of $W_\ell$ is absolutely irreducible. In what follows we construct an example of a weakly compatible system (for all $\ell \neq p$) of absolutely irreducible de Rham representations that does not come globally from an abelian variety over a number field. However, we do not know whether it comes locally from abelian varieties.

Let $p$ be a prime and $H$ be a definite quaternion algebra over $Q$ that is ramified exactly at $p$ and $\infty$. In particular, for each prime $\ell \neq p$ we have a $Q_\ell$-algebra isomorphism

$$H \otimes_{Q} Q_\ell \cong M_2(Q_\ell).$$
Let $g \geq 4$ be an even integer. According to Shimura [1963] (see also the case of Type III $(e_0 = 1)$ with $m = g/2$ in [Oort 1988, Table 8.1 on p. 498] and [Oort and Zarhin 1995, table on p. 23]), there exists a complex $g$-dimensional abelian variety $X$, whose endomorphism algebra $\text{End}^0(X)$ is isomorphic to $H$. The same arguments as above (related to $D$) prove that there exists a $g$-dimensional abelian variety $B$ over a certain number field $K$ such that all endomorphisms of $B$ are defined over $K$ and $\text{End}^0(B) \cong H$. In particular, $B$ is absolutely simple. By the theorem of Faltings, if $\ell$ is a prime then the $G_K$-module $V_\ell(B)$ is semisimple and

$$\text{End}_{G_K}(V_\ell(B)) = H \otimes_Q \mathbb{Q}_\ell.$$ 

In particular, if $\ell \neq p$ then $\text{End}_{G_K}(V_\ell(B)) \cong \mathbb{M}_2(\mathbb{Q}_\ell)$ and therefore there are two isomorphic $\mathbb{Q}_\ell[G_K]$-submodules $U_{1,\ell}(B)$ and $U_{2,\ell}(B)$ in $V_\ell(B)$ such that

$$V_\ell(B) = U_{1,\ell}(B) \oplus U_{2,\ell}(B) \cong U_{1,\ell}(B) \oplus U_{1,\ell}(B) \cong U_{2,\ell}(B) \oplus U_{2,\ell}(B).$$

If we denote by $U_\ell$ the $\mathbb{Q}_\ell[G_K]$-module $U_{1,\ell}(B)$ then $\dim_{\mathbb{Q}_\ell}(U_\ell) = g$ and we get an isomorphism of $\mathbb{Q}_\ell[G_K]$-modules

$$V_\ell(B) \cong U_\ell \oplus U_\ell.$$ 

Clearly, the submodule $U_\ell$ is semisimple and

$$\mathbb{M}_2(\mathbb{Q}_\ell) = H \otimes_Q \mathbb{Q}_\ell = \text{End}_{G_K}(V_\ell(B)) = \mathbb{M}_2(\text{End}_{G_K}(U_\ell)).$$

This implies that $\text{End}_{G_K}(U_\ell) = \mathbb{Q}_\ell$, i.e., the $\ell$-adic (sub)representation

$$G_K \rightarrow \text{Aut}_{\mathbb{Q}_\ell}(U_\ell) \cong \text{GL}_g(\mathbb{Q}_\ell)$$

is absolutely irreducible. Clearly, for each $\sigma \in G_K$ its characteristic polynomial with respect to the action on $V_\ell(B)$ is the square of its characteristic polynomial with respect to the action on $U_\ell$. This implies that if $v$ is a nonarchimedean place $v$ of $K$ where $B$ has good reduction then for all primes $\ell \neq p$ such that $v \nmid \ell$ the characteristic polynomial of the frobenius element at $v$ with respect to its action on $U_\ell$ has rational coefficients and does not depend on $\ell$. In other words, $U_\ell$ is a weakly compatible system of (absolutely irreducible) $\ell$-adic representations. As above, locally for each $v \mid \ell$ the $G_{K_v}$-module $V_\ell(B)$ is de Rham with Hodge weights 0 and $-1$ with weights $g$, which implies that $U_\ell$ is also de Rham with the same Hodge–Tate weights, whose multiplicities are $g/2$.

**Theorem 4.5.** The weakly compatible system of $g$-dimensional absolutely irreducible representations $\{U_\ell\}_{\ell \neq p}$ constructed above does not come globally from an abelian variety defined over the field $K$.

If $v \mid \ell$ then $U_\ell$ is locally a de Rham representation with Hodge–Tate weights 0 and $-1$, both of multiplicity $g/2$. 

Proof. We claim that none of \( U_\ell \) comes out from an abelian variety over \( K \). Indeed, if there is an abelian variety \( C \) over \( K \) such that the \( G_K \)-modules \( V_\ell(C) \) and \( U_\ell \) are isomorphic then \( \dim(C) = g/2 \) and the theorem of Faltings implies the existence of a nonzero homomorphism \( C \to B \), which contradicts the simplicity of \( g \)-dimensional \( B \).

\[ \square \]

5. Moduli of curves

The moduli space of smooth projective curves of genus \( g \) is denoted by \( \mathcal{M}_g \). It is also an orbifold and we will consider its fundamental group as such. For definitions see [Hain 2011]. It is defined over \( \mathbb{Q} \) and thus we can consider it over an arbitrary number field \( K \). As per our earlier conventions, \( \overline{\mathcal{M}}_g \) is the base change of \( \mathcal{M}_g \) to an algebraic closure of \( \mathbb{Q} \) and not a compactification.

Let \( X \) be a curve of genus \( g \) defined over \( K \). There is a map (an arithmetic analogue of the Dehn–Nielsen–Baer theorem, see [Matsumoto and Tamagawa 2000], in particular, Lemma 2.1) \( \rho : \pi_1(\mathcal{M}_g) \to \text{Out}(\pi_1(\overline{X})) \). This follows by considering the universal curve \( C_g \) of genus \( g \) together with the map \( C_g \to \mathcal{M}_g \), so \( X \) can be viewed as a fiber of this map. This gives rise to the fibration exact sequence

\[ 1 \to \pi_1(\overline{X}) \to \pi_1(C_g) \to \pi_1(\mathcal{M}_g) \to 1 \]

and the action of \( \pi_1(C_g) \) on \( \pi_1(\overline{X}) \) gives \( \rho \). Now, \( X \), viewed as a point on \( \mathcal{M}_g(K) \), gives a map \( \sigma_{\mathcal{M}_g/K}(X) : G_K \to \pi_1(\mathcal{M}_g) \). As pointed out in [Matsumoto and Tamagawa 2000], \( \rho \circ \sigma_{\mathcal{M}_g/K}(X) \) induces a map \( G_K \to \text{Out}(\pi_1(\overline{X})) \) which is none other than the map obtained from the exact sequence (1) by letting \( \pi_1(X) \) act on \( \pi_1(\overline{X}) \) by conjugation. Combining this with Theorem 2.1 (Mochizuki) gives:

**Theorem 5.1.** For any field \( K \) contained in a finite extension of a \( p \)-adic field, the section map \( \sigma_{\mathcal{M}_g/K} \) is injective.

The following result confirms a conjecture of Stoll [2007] if we assume that \( \sigma_{\mathcal{M}_g/K} \) surjects onto \( S_0(K, \mathcal{M}_g) \).

**Theorem 5.2.** Assume that \( \sigma_{\mathcal{M}_g/K}(\mathcal{M}_g(K)) = S_0(K, \mathcal{M}_g) \) for all \( g > 1 \) and all number fields \( K \). Then \( \sigma_{X/K}(X(K)) = S(K, X) \) for all smooth projective curves of genus at least two and all number fields \( K \).

**Proof.** For any algebraic curve \( X/K \) there is a nonconstant map \( X \to \mathcal{M}_g \) with image \( Y \), say, for some \( g \), defined over an extension \( L \) of \( K \), given by the Kodaira–Parshin construction. This gives a map \( \gamma : \pi_1(X \otimes L) \to \pi_1(\mathcal{M}_g \otimes L) \), over \( L \). Let \( s \in S(K, X) \), then \( \gamma \circ (s|_{G_L}) \in S_0(L, \mathcal{M}_g) \) and the assumption of the theorem yields that \( \gamma \circ (s|_{G_L}) = \sigma_{\mathcal{M}_g/L}(P) \), \( P \in \mathcal{M}_g(L) \). We can combine this with the injectivity of \( \sigma_{\mathcal{M}_g/K_v} \) (Mochizuki’s theorem) to deduce that in fact \( P \in Y(L_v) \cap \mathcal{M}_g(L) = Y(L) \). We can consider the pullback to \( X \) of the Galois orbit of \( P \), which gives us a zero.
dimensional scheme in $X$ having points locally everywhere and, moreover, being unobstructed by every abelian cover coming from an abelian cover of $X$. By the work of Stoll [2007, Proposition 5.2], we conclude that $X$ has a rational point corresponding to $s$. □

Acknowledgements

Voloch would like to thank J. Achter, D. Harari, E. Ozman, T. Schlank, and J. Starr for comments and information. He would also like to thank the Simons Foundation (grant #234591) and the Centre Bernoulli at EPFL for financial support.

Zarhin is grateful to Frans Oort, Ching-Li Chai and Jiangwei Xue for helpful discussions and to the Simons Foundation for financial and moral support (via grant #246625 to Yuri Zarkhin). Part of this work was done in May–June 2015 when he was visiting Department of Mathematics of the Weizmann Institute of Science (Rehovot, Israel). The final version of this paper was prepared in May–June 2016 when he was a visitor at the Max-Planck-Institut für Mathematik (Bonn, Germany). The hospitality and support of both institutes is gratefully acknowledged.

We are very grateful to the anonymous referees, whose careful readings and comments have greatly improved the readability of this paper. We would also like to thank W. Sawin for comments.

References


Communicated by Brian Conrad
Received 2015-07-14 Revised 2016-05-26 Accepted 2016-06-25

patrikis@math.utah.edu Department of Mathematics, University of Utah, 155 S 1400 E, Salt Lake City, UT 84112, United States

felipe.voloch@canterbury.ac.nz School of Mathematics and Statistics, University of Canterbury, Private Bag 4800, Christchurch 8140, New Zealand

voloch@math.utexas.edu Department of Mathematics, University of Texas, Austin, TX 78712, United States

zarhin@math.psu.edu Department of Mathematics, Pennsylvania State University, University Park, PA 16802, United States
On the local Tamagawa number conjecture for Tate motives over tamely ramified fields

Jay Daigle and Matthias Flach

The local Tamagawa number conjecture, which was first formulated by Fontaine and Perrin-Riou, expresses the compatibility of the (global) Tamagawa number conjecture on motivic L-functions with the functional equation. The local conjecture was proven for Tate motives over finite unramified extensions $K / \mathbb{Q}_p$ by Bloch and Kato. We use the theory of $(\phi, \Gamma)$-modules and a reciprocity law due to Cherbonnier and Colmez to provide a new proof in the case of unramified extensions, and to prove the conjecture for $\mathbb{Q}_p(2)$ over certain tamely ramified extensions.

1. Introduction

Let $K / \mathbb{Q}_p$ be a finite extension and $V$ a de Rham representation of $G_K := \text{Gal}(\overline{K} / K)$. The local Tamagawa number conjecture is a statement describing a certain $\mathbb{Q}_p$-basis of the determinant line $\text{det}_{\mathbb{Q}_p} R \Gamma(K, V)$ of (continuous) local Galois cohomology up to units in $\mathbb{Z}_p^\times$. It was first formulated by Fontaine and Perrin-Riou [1994, 4.5.4] as conjecture $C_{EP}$ and independently by Kato [1993, Conjecture 1.8] as the “local $\epsilon$-conjecture”. Both conjectures express compatibility of the (global) Tamagawa number conjecture on motivic L-functions with the functional equation. The fact that the local Tamagawa number conjecture is equivalent to this compatibility still constitutes its main interest. For example, the proof of the Tamagawa number conjecture for Dirichlet L-functions at integers $r \geq 2$ [Burns and Flach 2006] uses the conjecture at $1 - r$ and compatibility with the functional equation (no other more direct proof is known). Fukaya and Kato [2006] generalized [Kato 1993, Conjecture 1.8] to de Rham representations with coefficients in a possibly noncommutative $\mathbb{Q}_p$-algebra, and in fact to arbitrary $p$-adic families of local Galois representations.

In this paper we shall only consider Tate motives $V = \mathbb{Q}_p(r)$ with $r \geq 2$ (for the case $r = 1$ see [Bley and Cobbe 2016; Breuning 2004]). If $K / \mathbb{Q}_p$ is unramified the local Tamagawa number conjecture for $\mathbb{Q}_p(r)$ was first proven by Bloch and

MSC2010: primary 14F20; secondary 11G40, 18F10, 22A99.
Keywords: Tamagawa number conjecture.
Kato [1990] in their seminal paper on the global Tamagawa number conjecture, and has since been reproven by a number of authors (e.g., [Perrin-Riou 1994; Benois and Berger 2008]). These later proofs also cover the case where \( K/\mathbb{Q}_p \) is a cyclotomic extension, or more generally where \( V \) is an abelian de Rham representations of \( \text{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p) \) [Kato 1993, Theorem 4.1; Venjakob 2013]. All proofs have two main ingredients: Iwasawa theory and a “reciprocity law”. The latter is an explicit description of the exponential or dual exponential map for the de Rham representation \( V \), which however very often only holds in restricted situations (e.g., \( V \) ordinary or absolutely crystalline). The aim of this paper is to explore the application of the very general reciprocity law of Cherbonnier and Colmez [1999], which holds for arbitrary de Rham representations, to the local Tamagawa number conjecture for Tate motives.

In Section 2 we give a first somewhat explicit statement (Proposition 2) which is equivalent to the local Tamagawa conjecture for \( \mathbb{Q}_p(r) \) over an arbitrary Galois extension \( K/\mathbb{Q}_p \). We in fact work with the refined equivariant conjecture over the group ring \( \mathbb{Z}_p[\text{Gal}(K/\mathbb{Q}_p)] \), following Fukaya and Kato [2006]. In Section 3 we focus on the case where \( p \nmid [K: \mathbb{Q}_p] \). In Section 4 we state the reciprocity law of Cherbonnier and Colmez in the case of Tate motives. In Section 5 we show that it also can be used to give a proof of the unramified case (which however has many common ingredients with the existing proofs). Finally, in Section 6 we formulate our main result, Proposition 44, which is a fairly explicit statement equivalent to the equivariant local Tamagawa number conjecture for \( \mathbb{Q}_p(r) \) over \( K/\mathbb{Q}_p \) with \( p \nmid [K: \mathbb{Q}_p] \). We show that it can be used to prove some new cases; more specifically we have:

**Proposition 1.** Assume \( K/\mathbb{Q}_p \) is Galois of degree prime to \( p \) and with ramification degree \( e < p/4 \). Then the equivariant local Tamagawa number conjecture holds for \( V = \mathbb{Q}_p(2) \).

The only cases where the conjecture for tamely ramified fields was known previously are cyclotomic fields, i.e., where \( e \mid p - 1 \), and in this case one can allow arbitrary \( r \) [Perrin-Riou 1994; Benois and Berger 2008]. We believe many more cases can be proven with Proposition 44 and hope to return to this in a subsequent article.

### 2. The conjecture

Throughout this paper \( p \) denotes an odd prime. Let \( K/\mathbb{Q}_p \) be an arbitrary finite Galois extension with group \( G \) and \( r \geq 2 \). In this section we shall explicate the consequences of the local Tamagawa number conjecture of Fukaya and Kato [2006, Conjecture 3.4.3] for the triple

\[(A, T, \zeta) = (\mathbb{Z}_p[G], \text{Ind}_{G_k}^{\mathbb{Q}_p} \mathbb{Z}_p(1-r), \zeta).\]
Here $\zeta = (\zeta_p^n)_n \in \Gamma(\mathbb{Q}_p, \mathbb{Z}_p(1))$ is a compatible system of $p^n$-th roots of unity which we fix throughout this paper. The conjectures for a triple $(\Lambda, T, \zeta)$ and its dual $(\Lambda^o, T^*(1), \zeta)$ are equivalent. We find it advantageous to work with $\mathbb{Q}_p(1 - r)$ rather than $\mathbb{Q}_p(r)$ as in [Bloch and Kato 1990] since we are employing the Cherbonnier–Colmez reciprocity law [Cherbonnier and Colmez 1999] which describes the dual exponential map.

In order to give an idea what the conjecture is about, consider the Bloch–Kato exponential map [Bloch and Kato 1990]

$$\exp: K \to H^1(K, \mathbb{Q}_p(r)).$$

In a first approximation one may say that the local Tamagawa number conjecture describes the relation between the two $\mathbb{Z}_p$-lattices $\exp(O_K)$ and $\text{im}(H^1(K, \mathbb{Z}_p(r)))$ inside $H^1(K, \mathbb{Q}_p(r))$. Rather than giving a complete description of the relative position of these two lattices, the conjecture only specifies their relative volume, that is the class in $\mathbb{Q}_p^\times / \mathbb{Z}_p^\times$ which multiplies $\text{Det}_{\mathbb{Z}_p} \exp(O_K)$ to $\text{Det}_{\mathbb{Z}_p}(\text{im}(H^1(K, \mathbb{Z}_p(r))))$ inside the $\mathbb{Q}_p$-line $\text{Det}_{\mathbb{Q}_p} H^1(K, \mathbb{Q}_p(r))$. The equivariant form of the conjecture is a finer statement which arises by replacing determinants over $\mathbb{Z}_p$ by determinants over $\mathbb{Z}_p[G]$. If $G$ is abelian and $\text{im}(H^1(K, \mathbb{Z}_p(r)))$ is projective over $\mathbb{Z}_p[G]$, the conjecture thereby does specify the relative position of the two lattices in view of the fact that $H^1(K, \mathbb{Q}_p(r))$ is free of rank one over $\mathbb{Q}_p[G]$ and so coincides with its determinant. If $G$ is nonabelian, even though $H^1(K, \mathbb{Q}_p(r))$ remains free of rank one over $\mathbb{Q}_p[G]$, the conjecture is an identity in the algebraic $K$-group $K_1(\mathbb{Q}_p[G]))/K_1(\mathbb{Z}_p[G]))$ and is again quite a bit weaker than a full determination of the relative position of the two lattices.

Determinants in the sense of [Deligne 1987] (see also [Fukaya and Kato 2006, 1.2]) are only defined for modules of finite projective dimension, or more generally perfect complexes, and so the first step is to replace the $\mathbb{Z}_p$-lattice $\text{im}(H^1(K, \mathbb{Z}_p(r)))$ by the entire perfect complex $R\Gamma(K, \mathbb{Z}_p(r))$. There still is an isomorphism

$$R\Gamma(K, \mathbb{Z}_p(r)) \otimes_{\mathbb{Z}_p} \mathbb{Q}_p \cong R\Gamma(K, \mathbb{Q}_p(r)) \cong H^1(K, \mathbb{Q}_p(r))[-1]$$

(1)

since the groups $H^1(K, \mathbb{Z}_p(r))_\text{tor}$ and $H^2(K, \mathbb{Z}_p(r))$ are finite. If $K/\mathbb{Q}_p$ is Galois with group $G$ then $R\Gamma(K, \mathbb{Z}_p(r))$ is always a perfect complex of $\mathbb{Z}_p[G]$-modules whereas $\text{im}(H^1(K, \mathbb{Z}_p(r)))$ or $O_K$ need no longer have finite projective dimension over $\mathbb{Z}_p[G]$. A further simplification occurs if one does not try to compare $R\Gamma(K, \mathbb{Z}_p(r))$ to $\exp(O_K)$ directly. Instead one uses the “period isomorphism”

$$\text{per}: \mathbb{Q}_p \otimes_{\mathbb{Q}_p} K \cong \mathbb{Q}_p \otimes_{\mathbb{Q}_p} (\text{Ind}_{G_K}^{G_{\mathbb{Q}_p}} \mathbb{Q}_p) \cong \mathbb{Q}_p[G]$$

and tries to compare $\text{Det}_{\mathbb{Z}_p} R\Gamma(K, \mathbb{Z}_p(r))$ to a suitable lattice in this last space. The left-$\mathbb{Z}_p[G]$-module $\text{Ind}_{G_K}^{G_{\mathbb{Q}_p}} \mathbb{Z}_p$ is always free of rank one whereas $O_K$ need not be. After choosing an embedding $K \to \mathbb{Q}_p$ one gets an isomorphism $\psi: G_{\mathbb{Q}_p}/G_K \cong G$
and an isomorphism
\[ \text{Ind}_{G_p}^{G_{\mathbb{Q}_p}} \mathbb{Z}_p \cong \mathbb{Z}_p[G] \] (2)
so that the \( \mathbb{Z}_p[G] \)-linear left action of \( \gamma \in G_{\mathbb{Q}_p} \) is given by
\[ \mathbb{Z}_p[G] \ni x \mapsto x\psi(\gamma^{-1}). \] (3)
The period isomorphism is then given for \( x \in K \) by
\[ \text{per}(x) := \text{per}(1 \otimes x) = \sum_{g \in G} g(x) \cdot g^{-1} \in \overline{\mathbb{Q}}_p[G]. \]
The dual of \( \exp \) identifies with the dual exponential map
\[ \exp_{\mathbb{Q}_p}(r) : H^1(K, \mathbb{Q}_p(1-r)) \to K \]
by local Tate duality and the trace pairing on \( K \). Let \( \beta \in H^1(K, \mathbb{Z}_p(1-r)) \) be an element spanning a free \( \mathbb{Z}_p[G] \)-submodule and let \( C_{\beta} \) be the mapping cone of the ensuing map of perfect complexes of \( \mathbb{Z}_p[G] \)-modules
\[ (\mathbb{Z}_p[G] \cdot \beta)[-1] \to H^1(K, \mathbb{Z}_p(1-r))[−1] \to R\Gamma(K, \mathbb{Z}_p(1-r)). \]
Then \( C_{\beta} \) is a perfect complex of \( \mathbb{Z}_p[G] \)-modules with finite cohomology groups, i.e., such that \( C_{\beta} \otimes_{\mathbb{Z}_p} \mathbb{Q}_p \) is acyclic. It therefore represents a class \( [C_{\beta}] \) in the relative \( K \)-group \( K((\mathbb{Q}_p[G], \mathbb{Q}_p)) \) for which one has an exact sequence
\[ K_1(\mathbb{Z}_p[G]) \to K_1(\mathbb{Q}_p[G]) \to K_0(\mathbb{Z}_p[G], \mathbb{Q}_p) \to 0. \]
Hence we may also view \( [C_{\beta}] \) as an element in \( K_1(\mathbb{Q}_p[G]) / \text{im}(K_1(\mathbb{Z}_p[G])) \).
Extending scalars to \( \overline{\mathbb{Q}}_p \) we get an isomorphism of free rank-one \( \overline{\mathbb{Q}}_p[G] \)-modules
\[ H^1(K, \mathbb{Q}_p(1-r)) \otimes_{\mathbb{Q}_p} \overline{\mathbb{Q}}_p \xrightarrow{\exp^* \otimes \overline{\mathbb{Q}}_p} K \otimes_{\mathbb{Q}_p} \overline{\mathbb{Q}}_p \xrightarrow{\text{per}} \overline{\mathbb{Q}}_p[G] \]
sending the \( \overline{\mathbb{Q}}_p[G] \)-basis \( \beta \) to a unit \( \text{per}(\exp^*(\beta)) \in \overline{\mathbb{Q}}_p[G]^\times \). As such it has a class
\[ [\text{per}(\exp^*(\beta))] \in K_1(\overline{\mathbb{Q}}_p[G]) \]
via the natural projection map \( \overline{\mathbb{Q}}_p[G]^\times \to K_1(\overline{\mathbb{Q}}_p[G]) \) (recall that for any ring \( R \) we have maps \( R^\times \to \text{GL}(R) \to \text{GL}(R)^{ab} =: K_1(R) \)). In Section 2.2 below we shall define an \( \epsilon \)-factor \( \epsilon(K/\mathbb{Q}_p, 1-r) \in K_1(\overline{\mathbb{Q}}_p[G]) \) such that
\[ \epsilon(K/\mathbb{Q}_p, 1-r) \cdot [\text{per}(\exp^*(\beta))] \in K_1(\mathbb{Q}_p^{ur}[G]). \]
Let \( F \subset K \) denote the maximal unramified subfield, \( \Sigma = \text{Gal}(F/\mathbb{Q}_p) \) and \( \sigma \in \Sigma \) the (arithmetic) Frobenius automorphism. Then \( \mathbb{Q}_p(\Sigma) \) is canonically a direct factor of \( \mathbb{Q}_p[G] \) and \( \mathbb{Q}_p(\Sigma)^\times \cong K_1(\mathbb{Q}_p[\Sigma]) \) a direct factor of \( K_1(\mathbb{Q}_p[G]) \). For \( \alpha \in \mathbb{Q}_p(\Sigma)^\times \) we denote by \([\alpha]_F \) its class in \( K_1(\mathbb{Q}_p[G]) \). Finally, note that if \( R \) is a \( \mathbb{Q} \)-algebra then any nonzero rational number \( n \) has a class \([n] \in K_1(R) \) via \( \mathbb{Q}^\times \to R^\times \to K_1(R) \).
Proposition 2. Let $K/\mathbb{Q}_p$ be Galois with group $G$ and $r \geq 2$. The local Tamagawa number conjecture for the triple

$$(\Lambda, T, \zeta) = (\mathbb{Z}_p[G], \text{Ind}_{G_K}^{G_p} \mathbb{Z}_p(1 - r), \zeta).$$

is equivalent to the identity

$$[(r - 1)!] \cdot \epsilon(K/\mathbb{Q}_p, 1 - r) \cdot [\text{per}(\exp^*(\beta))] \cdot [C_{\beta}]^{-1} \cdot \left[ \frac{1 - p^r - 1 \sigma}{1 - p - r \sigma - 1} \right]_F = 1$$

in the group $K_1(\mathbb{Q}_p[G]) / \text{im}(K_1(\mathbb{Z}_p[G]))$.

Before we begin the proof of the proposition we explain what we mean by the local Tamagawa number conjecture for $(\mathbb{Z}_p[G], \text{Ind}_{G_K}^{G_p} \mathbb{Z}_p(1 - r), \zeta)$. The local Tamagawa number conjecture [Fukaya and Kato 2006, Conjecture 3.4.3] claims the existence of $\epsilon$-isomorphisms $\epsilon_{\Lambda, \zeta}(T)$ for all triples $(\Lambda, T, \zeta)$, where $\Lambda$ is a semilocal pro-$p$ ring satisfying a certain finiteness condition [Fukaya and Kato 2006, 1.4.1], $T$ a finitely generated projective $\Lambda$-module with continuous $G_{\mathbb{Q}_p}$-action and $\zeta$ a basis of $\Gamma(\mathbb{Q}_p, \mathbb{Z}_p(1))$, such that certain functorial properties hold. One of these properties [Fukaya and Kato 2006, Conjecture 3.4.3(v)] says that if $L := \Lambda \otimes_{\mathbb{Z}_p} \mathbb{Q}_p$ is a finite extension of $\mathbb{Q}_p$ and $V := T \otimes_{\mathbb{Z}_p} \mathbb{Q}_p$ is a de Rham representation, then

$$\tilde{L} \otimes_{\tilde{\Lambda}} \epsilon_{\Lambda, \zeta}(T) = \epsilon_{L, \zeta}(V),$$

where $\epsilon_{L, \zeta}(V)$ is the isomorphism in $C_{\tilde{L}}$ defined in [Fukaya and Kato 2006, 3.3]. Here, for any ring $R$, $C_R$ is the Picard category constructed in [Fukaya and Kato 2006, 1.2], equivalent to the category of virtual objects of [Deligne 1987], $S \otimes_R - : C_R \rightarrow C_S$ is the Picard functor induced by a ring homomorphism $R \rightarrow S$ and $\tilde{R} = W(\tilde{\mathbb{F}}_p) \otimes_{\mathbb{Z}_p} R$ for any $\mathbb{Z}_p$-algebra $R$. The construction of $\epsilon_{L, \zeta}(V)$ involves certain isomorphisms and exact sequences which we recall in the proof below. If $A$ is a finite dimensional semisimple $\mathbb{Q}_p$-algebra and $V$ an $A$-linear de Rham representation, those isomorphisms and exact sequences are in fact $A$-linear and therefore lead to an isomorphism $\epsilon_{A, \zeta}(V)$ in the category $C_{\tilde{A}}$. If $A := \Lambda \otimes_{\mathbb{Z}_p} \mathbb{Q}_p$ is a semisimple $\mathbb{Q}_p$-algebra and $V := T \otimes_{\mathbb{Z}_p} \mathbb{Q}_p$ is a de Rham representation, we say that the local Tamagawa number conjecture holds for the particular triple $(\Lambda, T, \zeta)$ if

$$\tilde{A} \otimes_{\tilde{\Lambda}} \epsilon_{\Lambda, \zeta}(T) = \epsilon_{A, \zeta}(V)$$

for some isomorphism $\epsilon_{A, \zeta}(T)$ in $C_{\tilde{A}}$.

Proof of Proposition 2. For a perfect complex of $\mathbb{Q}_p[G]$-modules $P$, we set

$$P^* = \text{Hom}_{\mathbb{Q}_p[G]}(P, \mathbb{Q}_p[G]).$$
which is a perfect complex of $\mathbb{Q}_p[G]^{\text{op}}$-modules. Fix $r \geq 2$ and set

$$V = \text{Ind}_{G_k}^{G_{\mathbb{Q}_p}} \mathbb{Q}_p(1 - r) \quad \text{and} \quad V^*(1) = \text{Ind}_{G_k}^{G_{\mathbb{Q}_p}} \mathbb{Q}_p(r),$$

which are free of rank one over $\mathbb{Q}_p[G]$ and $\mathbb{Q}_p[G]^{\text{op}}$, respectively. We recall the ingredients of the isomorphism $\theta_{\mathbb{Q}_p[G]}(V)$ of [Fukaya and Kato 2006, 3.3.2] (or rather of its generalization from field coefficients to semisimple coefficients). The element $\zeta$ determines an element $t = \log(\zeta)$ of $B_{dR}$. We have

$$D_{\text{cris}}(V) = F \cdot t^{r-1}, \quad D_{dR}(V)/D_{dR}^0(V) = 0,$$

$$D_{\text{cris}}(V^*(1)) = F \cdot t^{-r}, \quad D_{dR}(V^*(1))/D_{dR}^0(V^*(1)) = K,$$

$$C_f(\mathbb{Q}_p, V) : F \xrightarrow{1 - p^{-r} \sigma} F,$$

$$C_f(\mathbb{Q}_p, V^*(1)) : F \xrightarrow{(1 - p^{-r} \sigma, \subset)} F \oplus K,$$

and commutative diagrams

$$\begin{align*}
\text{Det}_{\mathbb{Q}_p[G]}(0) & \xrightarrow{\eta(\mathbb{Q}_p, V)} \text{Det}_{\mathbb{Q}_p[G]} C_f(\mathbb{Q}_p, V) \cdot \text{Det}_{\mathbb{Q}_p[G]} D_{dR}(V)/D_{dR}^0(V) \\
& \xrightarrow{[1 - p^{-r} \sigma]_{F}^{-1}} \text{Det}_{\mathbb{Q}_p[G]}(0) \cdot \text{Det}_{\mathbb{Q}_p[G]}(0)^{-1} \cdot \text{Det}_{\mathbb{Q}_p[G]}(0) \\
\text{Det}_{\mathbb{Q}_p[G]}(0) & \xrightarrow{\eta'(\mathbb{Q}_p, V)} \text{Det}_{\mathbb{Q}_p[G]} C_f(\mathbb{Q}_p, V^*(1))^* \\
& \xrightarrow{[1 - p^{-r} \sigma^{-1}]_{F}} \text{Det}_{\mathbb{Q}_p[G]}(0) \cdot \text{Det}_{\mathbb{Q}_p[G]}(K^*)^{-1} \cdot \text{Det}_{\mathbb{Q}_p[G]}(K^*) \\
\text{Det}_{\mathbb{Q}_p[G]} C_f(\mathbb{Q}_p, V^*(1))^* & \xrightarrow{\text{Det}_{\mathbb{Q}_p[G]} \Psi_f(\mathbb{Q}_p, V^*(1))^{*, -1}} \text{Det}_{\mathbb{Q}_p[G]}(C(\mathbb{Q}_p, V)/C_f(\mathbb{Q}_p, V)) \\
& \xrightarrow{c} \text{Det}_{\mathbb{Q}_p[G]}(K^*)^{-1} \\
\text{Det}_{\mathbb{Q}_p[G]}(K^*)^{-1} & \xrightarrow{\Psi'} \text{Det}_{\mathbb{Q}_p[G]} H^*(\mathbb{Q}_p, V)
\end{align*}$$

where the vertical maps $c$ are induced by passage to cohomology. The morphism $\Psi'$ is $(\text{Det}_{\mathbb{Q}_p[G]}^{-1})$ of the inverse of the isomorphism

$$H^1(\mathbb{Q}_p, V) \xrightarrow{T} H^1(\mathbb{Q}_p, V^*(1))^* \xrightarrow{\exp_{V^*(1)}^*} K^*,$$

where $T$ is the local Tate duality isomorphism. For the isomorphism

$$\theta_{\mathbb{Q}_p[G]}(V) = \eta(\mathbb{Q}_p, V) \cdot (\text{Det}_{\mathbb{Q}_p[G]} \Psi_f(\mathbb{Q}_p, V^*(1))^{*, -1} \circ \eta(\mathbb{Q}_p, V^*(1))^{*, -1})$$
we obtain a commutative diagram

\[
\begin{array}{ccc}
\text{Det}_{\mathbb{Q}_p[G]}(0) & \xrightarrow{\theta_{\mathbb{Q}_p[G]}(V)} & \text{Det}_{\mathbb{Q}_p[G]} C(\mathbb{Q}_p, V) \cdot \text{Det}_{\mathbb{Q}_p[G]} D_{dR}(V) \\
\left[\frac{1-p^{-r-1}}{1-p^{-t-1}}\right]_F & \Uparrow & \\
\text{Det}_{\mathbb{Q}_p[G]}(0) & \xrightarrow{\theta'} & \text{Det}_{\mathbb{Q}_p[G]} H^\bullet(\mathbb{Q}_p, V) \cdot \text{Det}_{\mathbb{Q}_p[G]}(K)
\end{array}
\]

where \(\theta'\) is induced by the dual exponential map

\[
H^1(\mathbb{Q}_p, V) \xrightarrow{\exp_{\ast_{\ast}(l)}} K.
\]

The isomorphism \(\Gamma_{\mathbb{Q}_p[G]}(V) \cdot \epsilon_{\mathbb{Q}_p[G], \xi, dR}(V)\) of [Fukaya and Kato 2006, 3.3.3] is the isomorphism

\[
[(−1)^r−1(r−1)!] \cdot \epsilon(\mathbb{Q}/\mathbb{Q}_p, 1−r) \cdot \text{Det}_{\mathbb{Q}_p[G]}(\text{per})
\]

and the isomorphism

\[
\epsilon_{\mathbb{Q}_p[G], \xi}(V) = \Gamma_{\mathbb{Q}_p[G]}(V) \cdot \epsilon_{\mathbb{Q}_p[G], \xi, dR}(V) \cdot \theta_{\mathbb{Q}_p[G]}(V)
\]

fits into a commutative diagram

\[
\begin{array}{ccc}
\text{Det}_{\mathbb{Q}_p[G]}(0) & \xrightarrow{\epsilon_{\mathbb{Q}_p[G], \xi}(V)} & \mathbb{Q}_p^ur[G] \otimes_{\mathbb{Q}_p[G]} (\text{Det}_{\mathbb{Q}_p[G]} R\Gamma(K, \mathbb{Q}_p(1−r)) \cdot \text{Det}_{\mathbb{Q}_p[G]}(V)) \\
\left[\frac{1-p^{-r-1}}{1-p^{-t-1}}\right]_F & \Uparrow & \\
\text{Det}_{\mathbb{Q}_p[G]}(0) & \xrightarrow{\theta''} & \mathbb{Q}_p^ur[G] \otimes_{\mathbb{Q}_p[G]} (\text{Det}_{\mathbb{Q}_p[G]}^{-1} H^1(K, \mathbb{Q}_p(1−r)) \cdot \text{Det}_{\mathbb{Q}_p[G]}(\mathbb{Q}_p[G]))
\end{array}
\]

where

\[
\theta'' = [(−1)^r−1(r−1)!] \cdot \epsilon(\mathbb{Q}/\mathbb{Q}_p, 1−r) \cdot \text{Det}_{\mathbb{Q}_p[G]}(\text{per}) \cdot \theta'
\]

and \(c\) involves passage to cohomology as well as our identification \(V \cong \mathbb{Q}_p[G]\) chosen above. Now passage to cohomology is also the scalar extension of the isomorphism

\[
\text{Det}_{\mathbb{Z}_p[G]}^{-1}(\mathbb{Z}_p[G] \cdot \beta) \cdot \text{Det}_{\mathbb{Z}_p[G]}(C_\beta) \cong \text{Det}_{\mathbb{Z}_p[G]} R\Gamma(K, \mathbb{Z}_p(1−r))
\]

induced by the short exact sequence of perfect complexes of \(\mathbb{Z}_p[G]\)-modules

\[
0 \rightarrow R\Gamma(K, \mathbb{Z}_p(1−r)) \rightarrow C_\beta \rightarrow \mathbb{Z}_p[G] \cdot \beta \rightarrow 0
\]

combined with the acyclicity isomorphism

\[
\text{can} : \text{Det}_{\mathbb{Q}_p[G]}(0) \cong \text{Det}_{\mathbb{Q}_p[G]}(C_\beta, \mathbb{Q}_p).
\]
Since the class of $C_\beta$ in $K_0(\mathbb{Z}_p[G])$ vanishes, we can choose an isomorphism
\[ a : \text{Det}_{\mathbb{Z}_p[G]}(0) \cong \text{Det}_{\mathbb{Z}_p[G]}(C_\beta), \]
which leads to another isomorphism
\[ c' : \text{Det}_{\mathbb{Z}_p[G]}(\mathbb{Z}_p[G] \cdot \beta) \cong \text{Det}_{\mathbb{Z}_p[G]}(K, \mathbb{Z}_p(1-r)) \]
defined over $\mathbb{Z}_p[G]$. Setting
\[ \lambda := (c'_{\mathbb{Q}_p})^{-1} \in \text{Aut}(\text{Det}_{\mathbb{Q}_p[G]}^{-1}(K, \mathbb{Q}_p(1-r))) = K_1(\mathbb{Q}_p[G]), \]
we obtain a commutative diagram
\[
\begin{array}{ccc}
\text{Det}_{\mathbb{Q}_p[G], \zeta}(V) & \xrightarrow{\epsilon_{\mathbb{Q}_p[G], \zeta}(V)} & \mathbb{Q}_p[G] \otimes_{\mathbb{Q}_p[G]} (\text{Det}_{\mathbb{Q}_p[G]}(K, \mathbb{Q}_p(1-r)) \cdot \text{Det}_{\mathbb{Q}_p[G]}(V)) \\
\text{Det}_{\mathbb{Z}_p[G]}(0) & \xrightarrow{\theta''} & \mathbb{Z}_p[G] \otimes_{\mathbb{Z}_p[G]} (\text{Det}_{\mathbb{Z}_p[G]}^{-1}(K, \mathbb{Q}_p(1-r)) \cdot \text{Det}_{\mathbb{Z}_p[G]}(\mathbb{Q}_p[G])) \\
\end{array}
\]
where
\[ \theta'' = \lambda \circ \theta'' = \lambda \cdot [(−1)^{t−1}(r−1)!] \cdot \epsilon(K/\mathbb{Q}_p, 1−r) \cdot \text{Det}_{\mathbb{Q}_p[G]}(\text{per}) \cdot \theta'. \]
The local Tamagawa number conjecture claims that $\epsilon_{\mathbb{Q}_p[G], \zeta}(V)$ is induced by an isomorphism
\[
\begin{array}{ccc}
\text{Det}_{\mathbb{Z}_p[G]}(0) & \xrightarrow{\epsilon_{\mathbb{Z}_p[G], \zeta}(T)} & \mathbb{Z}_p[G] \otimes_{\mathbb{Z}_p[G]} (\text{Det}_{\mathbb{Z}_p[G]}(K, \mathbb{Z}_p(1-r)) \cdot \text{Det}_{\mathbb{Z}_p[G]}(T)) \\
\end{array}
\]
and this will be the case if and only if
\[ \theta''' := \theta'' \cdot \left[ \frac{1−\sigma}{1−p^{r−1}\sigma} \right]_F \]
is induced by an isomorphism
\[
\begin{array}{ccc}
\text{Det}_{\mathbb{Z}_p[G]}(0) & \xrightarrow{\theta'''_{\mathbb{Z}_p[G]}} & \mathbb{Z}_p[G] \otimes_{\mathbb{Z}_p[G]} (\text{Det}_{\mathbb{Z}_p[G]}^{-1}(\mathbb{Z}_p[G] \cdot \beta) \cdot \text{Det}_{\mathbb{Z}_p[G]}(\mathbb{Z}_p[G])). \\
\end{array}
\]
The isomorphism of $\mathbb{Q}_p[G]$-modules
\[ \tau : H^1(K, \mathbb{Q}_p(1−r)) \otimes_{\mathbb{Q}_p} \mathbb{Q}_p \xrightarrow{\text{exp}^* \otimes \mathbb{Q}_p} K \otimes_{\mathbb{Q}_p} \mathbb{Q}_p \xrightarrow{\text{per}} \mathbb{Q}_p[G] \xrightarrow{\text{per}(\exp^* (\beta))^{-1}} \mathbb{Q}_p[G] \]
is clearly induced by an isomorphism of $\mathbb{Z}_p[G]$-modules
\[ \tau_{\mathbb{Z}_p[G]} : \mathbb{Z}_p[G] \cdot \beta \sim \mathbb{Z}_p[G] \]
and we have
\[
\theta^{iv} = \left[ \frac{1 - p^{-r-1}\sigma}{1 - p^{-r}\sigma^{-1}} \right]_F \cdot \lambda \cdot [(-1)^{r-1}(r - 1)!] 
\cdot \epsilon(K/\mathbb{Q}_p, 1 - r) \cdot [\text{per}(\exp^*(\beta))] \cdot \text{Det}_{\mathbb{Q}_p[G]}(\tau).
\]

Hence \( \theta^{iv} \) is induced by an isomorphism \( \theta^{iv}_{\mathbb{Z}_p[G_i]} \) if and only if the class in \( K_1(\mathbb{Q}_p[G]) \) of
\[
\left[ \frac{1 - p^{-r-1}\sigma}{1 - p^{-r}\sigma^{-1}} \right]_F \cdot \lambda \circ [(-1)^{r-1}(r - 1)!] \cdot \epsilon(K/\mathbb{Q}_p, 1 - r) \cdot [\text{per}(\exp^*(\beta))]
\]
lies in \( K_1(\mathbb{Z}_p^{ur}[G]) \). Now note that \([(-1)] \in K_1(\mathbb{Z}) \subset K_1(\mathbb{Z}_p^{ur}[G])\) and that \( \lambda = [C_\beta]^{-1} \), so we do indeed obtain identity (4). In order to see this last identity, note that we have
\[
\lambda^{-1} = a^{-1} \cdot \text{can}
\]
and that \( a^{-1} \cdot \text{can} \in K_1(\mathbb{Q}_p^{ur}[G]) \) is a lift of \([C_\beta] \in K_0(\mathbb{Z}_p^{ur}[G], \mathbb{Q}_p)\) according to the conventions of [Fukaya and Kato 2006, 1.3.8, Theorem 1.3.15(ii)].

\[\square\]

2.1. Description of \( K_1 \). For any finite group \( G \) we have the Wedderburn decomposition
\[
\mathbb{Q}_p[G] \cong \prod_{\chi \in \hat{G}} M_{d_\chi}(\mathbb{Q}_p),
\]
where \( \hat{G} \) is the set of irreducible \( \mathbb{Q}_p \)-valued characters of \( G \) and \( d_\chi = \chi(1) \) is the degree of \( \chi \). Hence there is a corresponding decomposition
\[
K_1(\mathbb{Q}_p[G]) \cong \prod_{\chi \in \hat{G}} K_1(M_{d_\chi}(\mathbb{Q}_p)) \cong \prod_{\chi \in \hat{G}} \mathbb{Q}_p^\times,
\]
which allows one to think of \( K_1(\mathbb{Q}_p[G]) \) as a collection of nonzero \( p \)-adic numbers indexed by \( \hat{G} \). Note here that for any ring \( R \) one has \( K_1(M_d(R)) = K_1(R) \) and for a commutative semilocal ring \( R \) one has \( K_1(R) = R^\times \).

If \( p \nmid |G| \) then all characters \( \chi \in \hat{G} \) take values in \( \mathbb{Z}_p^{ur} \), the Wedderburn decomposition is already defined over \( \mathbb{Z}_p^{ur} \) and so is the decomposition of \( K_1 \). One has
\[
K_1(\mathbb{Z}_p^{ur}[G]) \cong \prod_{\chi \in \hat{G}} K_1(M_{d_\chi}(\mathbb{Z}_p^{ur})) \cong \prod_{\chi \in \hat{G}} \mathbb{Z}_p^{ur,\times}
\]
and
\[
K_1(\mathbb{Q}_p^{ur}[G]) / \text{im}(K_1(\mathbb{Z}_p^{ur}[G])) \cong \prod_{\chi \in \hat{G}} \mathbb{Q}_p^{ur,\times} / \mathbb{Z}_p^{ur,\times} \cong \prod_{\chi \in \hat{G}} p^{\mathbb{Z}},
\]
which allows one to think of elements in \( K_1(\mathbb{Q}_p^{ur}[G]) / \text{im}(K_1(\mathbb{Z}_p^{ur}[G])) \) as a collection of integers (\( p \)-adic valuations) indexed by \( \hat{G} \).
2.2. Definition of the $\epsilon$-factor. If $L$ is a local field, $E$ an algebraically closed field of characteristic 0 with the discrete topology, $\mu_L$ a Haar measure on the additive group of $L$ with values in $E$, $\psi_L : L \to E^\times$ a continuous character, the theory of Langlands–Deligne [Deligne 1973] associates to each continuous representation $r$ of the Weil group $W_L$ over $E$ an $\epsilon$-factor $\epsilon(r, \psi_L, \mu_L) \in E^\times$.

We shall take $E = \mathbb{Q}_p$ and always fix $\mu_L$ and $\psi_L$ so that $\mu_L(O_L) = 1$ and $\psi_L = \psi_{\mathbb{Q}_p} \circ \text{Tr}_{L/\mathbb{Q}_p}$ where $\psi_{\mathbb{Q}_p}(p^{-n}) = \zeta_{p^n}$ for our fixed $\zeta = (\zeta_{p^n})_n \in \Gamma(\mathbb{Q}_p, \mathbb{Z}_p(1))$. Setting $\epsilon(r) := \epsilon(r, \psi_L, \mu_L) \in E^\times$ and leaving the dependence on $\zeta$ implicit, we have the following properties (see also [Benois and Berger 2008] for a review, [Fukaya and Kato 2006] only reviews the case $L = \mathbb{Q}_p$). Let $\pi$ be a uniformizer of $O_L$, $\delta_L$ the exponent of the different of $L/\mathbb{Q}_p$ and $q = |O_L/\pi|$.

(a) If $r : W_L \to E^\times$ is a homomorphism, set $r^\#: L \times \text{rec} \to W_{ab}^L \to E^\times$ where rec is normalized as in [Deligne 1973, (2.3)] and sends a uniformizer to a geometric Frobenius automorphism in $W_{ab}^L$. Then we have

$$
\epsilon(r) = \begin{cases} 
q^{\delta_L} & \text{if } c = 0, \\
q^{\delta_L} r^\#_\pi(\pi^c + \delta_L) \tau(r^\#, \psi_\pi) & \text{if } c > 0,
\end{cases}
$$

where $c \in \mathbb{Z}$ is the conductor of $r$ and

$$
\tau(r^\#, \psi_\pi) = \sum_{u \in (O_L/\pi^c)^\times} r^\#_\pi^{-1}(u) \psi_\pi(u)
$$

is the Gauss sum associated to the restriction of $r^\#$ to $(O_L/(\pi^c))^\times$ and the additive character

$$
u \mapsto \psi_\pi(u) := \psi_K(\pi^{-\delta_L-c}u)
$$

of $O_L/(\pi^c)$.

(b) If $L/K$ is unramified then $\epsilon(r) = \epsilon(\text{Ind}_{W_K}^{W_L} r)$ for any representation $r$ of $W_L$.

(c) If $r(\alpha)$ is the twist of $r$ with the unramified character with Frob-$L$-eigenvalue $\alpha \in E^\times$, and $c(r) \in \mathbb{Z}$ is the conductor of $r$, then

$$
\epsilon(r(\alpha)) = \alpha^{-c(r) - \dim_E(r)\delta_L} \epsilon(r).
$$

Here Frob$_L$ denotes the usual (arithmetic) Frobenius automorphism.

For a potentially semistable representation $V$ of $G_{\mathbb{Q}_p}$ one first forms $D_{\text{pst}}(V)$, a finite dimensional $\hat{\mathbb{Q}}_{p\text{ur}}$-vector space of dimension $\dim_{\mathbb{Q}_p} V$ with an action of $G_{\mathbb{Q}_p}$, semilinear with respect to the natural action of $G_{\mathbb{Q}_p}$ on $\hat{\mathbb{Q}}_{p\text{ur}}$ and discrete on the inertia subgroup. Moreover, $D_{\text{pst}}(V)$ has a Frob-semilinear automorphism $\varphi$. The
associated linear representation $r_V$ of $W_{Q_p}$ over $E = \hat{Q}_{Q_p}^{ur}$ is the space $D_{pst}(V)$ with action

$$r_V(w)(d) = \iota(w)\varphi^{-v(w)}(d),$$

where $\iota : W_{Q_p} \to G_{Q_p}$ is the inclusion and $v(w) \in \mathbb{Z}$ is such that $\text{Frob}^{v(w)}$ is the image of $w$ in $G_{\mathbb{F}_p}$.

From now on we are interested in $V = (\text{Ind}_{G_{Q_p}}^{G_{\hat{Q}_p}}(1 - r))$. Here one has

$$D_{pst}(V) = (\text{Ind}_{G_{Q_p}}^{G_{\hat{Q}_p}}(\hat{Q}_{Q_p}^{ur}) \cdot t^{r-1}, \quad r_V = (\text{Ind}_{W_{Q_p}}^{W_{\hat{Q}_p}}(\hat{Q}_{Q_p}^{ur})(p^{1-r})),

and we notice that $r_V$ is the scalar extension from $\hat{Q}_{Q_p}^{ur}$ to $\hat{Q}_{Q_p}^{ur}$ of the representation $(\text{Ind}_{W_{Q_p}}^{W_{\hat{Q}_p}} Q_{Q_p}^{ur})(p^{1-r})$. So completion of $Q_{Q_p}^{ur}$ is not needed in this example. Associated to $r_V \otimes Q_{Q_p}^{ur} \hat{Q}_{Q_p}$ is an $\epsilon$-factor in $\epsilon(r_V) \in K_1(\overline{Q}_p)$. However, as explained above before (3), $r_V$ carries a left action of $Q_{Q_p}^{ur}[G]$ commuting with the left $W_{Q_p}$-action, so we will actually be able to associate to $r_V \otimes Q_{Q_p}^{ur} \hat{Q}_{Q_p} \epsilon$ a refined $\epsilon$-factor

$$\epsilon(K/Q, 1 - r) \in K_1(\overline{Q}_p[G]).$$

For each $\chi \in \hat{G}$ define a representation $r_\chi$ of $W_{Q_p}$ over $E = \overline{Q}_p$ by

$$W_{Q_p} \xrightarrow{\iota} G_{Q_p} \xrightarrow{\psi} G \xrightarrow{\rho_\chi} \text{GL}_{d_\chi}(E),$$

where $\rho_\chi : G \to \text{GL}_{d_\chi}(E)$ is a homomorphism realizing $\chi$. Let $E_{d_\chi}$ be the space of row vectors on which $G$ acts on the right via $\rho_\chi$ and define another representation of $W_{Q_p}$ over $E = \overline{Q}_p$

$$r_{V, \chi} = E_{d_\chi} \otimes Q_{Q_p}^{ur}[G] r_V = E_{d_\chi} \otimes Q_{Q_p}^{ur}[G] (\text{Ind}_{W_{Q_p}}^{W_{\hat{Q}_p}} Q_{Q_p}^{ur})(p^{1-r}) \cong E_{d_\chi}.$$

By (3), the left $W_{Q_p}$-action on this last space is given by the contragredient $\iota \rho_\chi(\psi(g))^{-1}$ of $r_\chi$, twisted by the unramified character with eigenvalue $p^{1-r}$. So we have

$$r_{V, \chi} \cong r_{\hat{\chi}}(p^{1-r}),$$

where $\hat{\chi}$ is the contragredient character of $\chi$. We view the collection

$$\epsilon(K/Q, 1 - r) := (\epsilon(r_{V, \chi}))_{\chi \in \hat{G}} = (\epsilon(r_{\hat{\chi}}) p^{(r-1)e(r_{\hat{\chi}})})_{\chi \in \hat{G}}$$

as an element of $K_1(\overline{Q}_p[G])$ in the description (5).

3. The conjecture in the case $p \nmid |G|$  

In this section and for most of the rest of the paper we assume that $p$ does not divide $|G| = [K : Q_p]$. In particular $K/Q_p$ is tamely ramified with maximal unramified subfield $F$. Although our methods probably extend to an arbitrary tamely ramified extension $K/Q_p$ (i.e., where $p$ is allowed to divide $[F : Q_p]$) this would add an
extra layer of notational complexity which we have preferred to avoid. The group
\( G = \text{Gal}(K/\mathbb{Q}_p) \) is an extension of two cyclic groups
\[ \Sigma := \text{Gal}(F/\mathbb{Q}_p) \cong \mathbb{Z}/f\mathbb{Z}, \]
\[ \Delta := \text{Gal}(K/F) \cong \mathbb{Z}/e\mathbb{Z}, \]
where the action of \( \sigma \in \Sigma \) on \( \Delta \) is given by \( \delta \mapsto \delta^p \) and we have \( e \mid p^f - 1 \). By
Kummer theory \( K = F(\sqrt[p^0]{p}) \), where \( p_0 \in (F^\times/(F^\times)^e)^\Sigma \) has order \( e \). We can
and will assume that \( p_0 \) has \( p \)-adic valuation one, and in fact that \( p_0 = \lambda \cdot p \) with \( \lambda \in \mu_F \). Writing \( p_0 = \lambda' \cdot p'_0 \) with \( p'_0 \in \mathbb{Q}_p \) we see that \( K \) is contained in \( F'/(\sqrt[p'_0]{p}) \),
where \( F' := F(\sqrt[p'_0]{p}) \) is unramified over \( \mathbb{Q}_p \) and \( p'_0 \) is any choice of element in
\( \mu_{\mathbb{Q}_p} \cdot p = \mu_{p-1} \cdot p \). Since for the purpose of proving the local Tamagawa number
conjecture we can always enlarge \( K \), we may and will assume that
\[ K = F(\sqrt[p^0]{p}), \quad p_0 \in \mu_{p-1} \cdot p \subseteq \mathbb{Q}_p. \]
We then have
\[ G = \text{Gal}(K/\mathbb{Q}_p) \cong \Sigma \rtimes \Delta \]
since \( \text{Gal}(K/\mathbb{Q}_p(\sqrt[p^0]{p})) \) is a complement of \( \Delta \). If \( (e, p - 1) = 1 \), then the fields
\( K = F(\sqrt[p^0]{p}) \) for \( p_0 \in \mu_{p-1} \cdot p \) are all isomorphic; in fact any Galois extension
\( K/\mathbb{Q}_p \) with invariants \( e \) and \( f \) is then isomorphic to the field \( F(\sqrt[p]{p}) \).

The choice of \( p_0 \) (in fact just the valuation of \( p_0 \)) determines a character
\[ \eta_0 : \Delta \to \mu_e \subset F^\times \subset \mathbb{Q}_p^{ur, \times} \subset \overline{\mathbb{Q}}_p^{ur, \times} \quad (10) \]
by the usual formula \( \delta(\sqrt[p^0]{p}) = \eta_0(\delta) \cdot \sqrt[p^0]{p} \). Let
\[ \eta : \Delta \to F^\times \]
be any character of \( \Delta \) and
\[ \Sigma_\eta := \{ g \in \Sigma \mid \eta(g\delta g^{-1}) = \eta(\delta) \quad \text{for all} \quad \delta \in \Delta \} \]
the stabilizer of \( \eta \). Then for any character \( \eta' : \Sigma_\eta \to \mathbb{Q}_p^{ur, \times} \) we obtain a character
\[ \eta' \eta : G_\eta := \Sigma_\eta \rtimes \Delta \to \mathbb{Q}_p^{ur, \times} \]
and an induced character
\[ \chi := \text{Ind}^{G}_{G_\eta}(\eta' \eta) \]
of \( G \). By [Lang 2002, Exercise XVIII.7], all irreducible characters of \( G \) are obtained
by this construction, and in fact each \( \chi \in \hat{G} \) is parametrized by a unique pair \( ([\eta], \eta') \)
where \( [\eta] \) denotes the \( \Sigma \)-orbit of \( \eta \). The degree of \( \chi \) is given by
\[ d_\chi = \chi(1) = f_\eta := [\Sigma : \Sigma_\eta] = [F_\eta : \mathbb{Q}_p], \quad (11) \]
where \( F_\eta \subseteq F \) is the fixed field of \( \Sigma_\eta \).
We have
\[ r_\chi = \text{Ind}_{W_{\mathbb{Q}_p}}^{W_{F_\eta}}(r_{\eta'}\eta), \]
where \( r_\chi \) and \( r_{\eta'}\eta \) are the representations of \( W_{\mathbb{Q}_p} \) and \( W_{F_\eta} \), respectively, defined as in (8). By [Serre 1979, Chapter VI, Corollary to Proposition 4] we have
\[ c(r_\chi) = f_\eta c(r_\eta) = \begin{cases} 0, & \eta = 1, \\ f_\eta, & \eta \neq 1. \end{cases} \]
Using (b), (c) and (a) of Section 2.2 we have
\[ \epsilon(r_\chi) = \epsilon(r_{\eta'}\eta) \]
\[ = \begin{cases} 1, & \eta = 1, \\ \epsilon(r_\eta) r_{\eta'}(\text{Frob}_{F_\eta})^{-c(r_\eta)} = \eta(\text{rec}(p)) \tau(r_{\eta'}, \psi_p) \eta'(\sigma f_\eta)^{-1}, & \eta \neq 1. \end{cases} \]

3.1. Gauss sums. If \( k_\eta \) denotes the residue field of \( F_\eta \), we have a canonical character
\[ \omega : k_\eta^\times \hookrightarrow \mu_{p f_\eta - 1} \subseteq F_\eta^\times \subseteq K^\times \subseteq \overline{\mathbb{Q}}_p^\times, \]
where the first arrow is reduction mod \( p \). On the other hand we have our character
\[ r_{\eta', \sharp} : F_\eta^\times \xrightarrow{\text{rec}} W_{F_\eta}^{\text{ab}} \xrightarrow{i} G_{F_\eta}^{\text{ab}} \xrightarrow{\psi} G_\eta^{\text{ab}} \xrightarrow{\eta} \overline{\mathbb{Q}}_p^\times \]
of order dividing \( e \). So there exists a unique \( m_\eta \in \mathbb{Z}/e\mathbb{Z} \) such that
\[ r_{\eta', \sharp} |_{\mu_{p f_\eta - 1}} = \omega^{m_\eta (p f_\eta - 1)/e} \]
and formula (7) gives
\[ \tau(r_{\eta', \sharp}, \psi_p) = \tau(\omega^{-m_\eta (p f_\eta - 1)/e}), \]
where
\[ \tau(\omega^{-i}) := \sum_{a \in k_\eta^\times} \omega(a)^{-i} \xi_p^{\text{Tr}_{k_\eta/F_\eta}(a)} \]
is a Gauss sum associated to the finite field \( k_\eta \). The \( p \)-adic valuation of these sums is known:

**Lemma 3** [Washington 1997, Proposition 6.13 and Lemma 6.14]. For \( 0 \leq i \leq p f_\eta - 1 \), let \( i = i_0 + pi_1 + p^2i_2 + \cdots + i_{f_\eta - 1}p^{f_\eta - 1} \) be the \( p \)-adic expansion with digits \( 0 \leq i_j \leq p - 1 \). Then
\[ v_p(\tau(\omega^{-i})) = \frac{i_0 + i_1 + \cdots + i_{f_\eta - 1}}{p - 1} = \sum_{j=0}^{f_\eta - 1} \left\lfloor \frac{ip^j}{p f_\eta - 1} \right\rfloor, \]
where \( v_p : \overline{\mathbb{Q}}_p^\times \to \mathbb{Q} \) is the \( p \)-adic valuation on \( \overline{\mathbb{Q}}_p \), normalized by \( v_p(p) = 1 \) and \( 0 \leq \langle x \rangle < 1 \) is the fractional part of the real number \( x \).
Corollary 4. For all $\eta \in \hat{\Delta}$ we have

$$v_p(\tau(r_{\eta, \sharp}, \psi_p)) = \sum_{j=0}^{f_{\eta}/p-1} \left( \frac{m_{\eta} p^j}{e} \right).$$

After this interlude on Gauss sums we now prove a statement about periods of certain specific elements in $K$ which will eliminate any further reference to $\epsilon$-factors in the proof of Equation (4).

Proposition 5. Let $K/\mathbb{Q}_p$ be Galois with group $G$ of order prime to $p$. Then any fractional $\mathcal{O}_K$-ideal is a free $\mathbb{Z}_p[G]$-module of rank one and

$$(\epsilon(r_{\bar{\chi}}))_{\chi \in \hat{G}} \cdot [\text{per}(b)] \in \text{im}(K_1(\mathbb{Z}_p[G]))$$

for any $\mathbb{Z}_p[G]$-basis $b$ of the inverse different $(\sqrt[p]{p_0})^{-\delta_G} \mathcal{O}_K = (\sqrt[p]{p_0})^{-1} \mathcal{O}_K$.

Proof. This is a classical result in Galois module theory which can be found in [Fröhlich 1976] but rather than trying to match our notation to that paper we go through the main computations again. In this proof $\sigma$ will temporarily denote a generic element of $G$ rather than the Frobenius.

The image of $[\text{per}(b)]$ in the $\chi$-component of the decomposition (5) is the $(d_\chi \times d_\chi)$-determinant

$$[\text{per}(b)]_\chi := \det \rho_\chi \left( \sum_{g \in G} g(b) \cdot g^{-1} \right) = \det \sum_{g \in G} g(b) \rho_\chi(g)^{-1} \in \overline{\mathbb{Q}}^\times.$$

This character function is traditionally called a resolvent. With notation as above, $(\sqrt[p]{p_0})^{-1} \mathcal{O}_K$ is a free $\mathbb{Z}_p[G_\eta]$-module with basis $\sigma(b)$, where $\sigma \in G_\eta \setminus G \cong \Sigma_\eta \setminus \Sigma$ runs through a set of right coset representatives. The image of this basis under the period map is

$$\text{per}(\sigma(b)) = \sum_{g \in G} g \sigma(b) \cdot g^{-1} = \sum_{\tau \in \Sigma_\eta \setminus \Sigma} \left( \sum_{g \in G_\eta} \tau^{-1} g \sigma(b) \cdot g^{-1} \right) \tau$$

and if $\chi = \text{Ind}^G_{G_\eta}(\chi')$ is an induced character we have by [Fröhlich 1976, (5.15)]

$$\rho_\chi \left( \sum_{g \in \hat{G}} g(b) \cdot g^{-1} \right) = \left( \sum_{g \in G_\eta} \tau^{-1} g \sigma(b) \cdot \rho_{\chi'}(g)^{-1} \right)_{\sigma, \tau}. $$

In our case of interest $\chi' = \eta' \eta$ is a one-dimensional character. Write

$$b = \xi \cdot x,$$
where $x$ is an $\mathcal{O}_F[\Delta]$-basis of $(\sqrt[6]{\mathcal{O}_F})^{(e-1)}\mathcal{O}_K$ fixed by $\Sigma$ and $\xi$ a $\mathbb{Z}_p[\Sigma]$-basis of $\mathcal{O}_F$. Then writing $g = \delta\sigma'$ with $\delta \in \Delta$ and $\sigma' \in \Sigma_{\eta}$ this matrix becomes

$$\left( \sum_{\sigma' \in \Sigma_{\eta}} \tau^{-1}\sigma'\sigma(\xi)\eta'(\sigma')^{-1} \sum_{\delta \in \Delta} \tau^{-1}\delta(x) \cdot \eta(\delta)^{-1} \right)_{\sigma, \tau}$$

and its determinant is

$$\det \left( \sum_{\sigma' \in \Sigma_{\eta}} \tau^{-1}\sigma'\sigma(\xi)\eta'(\sigma')^{-1} \right)_{\sigma, \tau} \cdot \prod_{\tau \in \Sigma_{\eta}\backslash \Sigma} \sum_{\delta \in \Delta} \tau^{-1}\delta \cdot \eta(\delta)^{-1}.$$  

The first determinant is a group determinant [Washington 1997, Lemma 5.26] for the group $\Sigma_{\eta}\backslash \Sigma$ and equals

$$\xi_{\eta'} := \prod_{\kappa \in (\Sigma_{\eta}\backslash \Sigma)^{\wedge}} \sum_{\kappa \in \Sigma_{\eta}\backslash \Sigma} \left( \sum_{\sigma' \in \Sigma_{\eta}} \sigma'\sigma(\xi)\eta'(\sigma')^{-1} \right) \kappa(\sigma)^{-1} = \prod_{\kappa} \sum_{\sigma \in \Sigma} \sigma(\xi)\kappa(\sigma)^{-1},$$

where this last product is over all characters $\kappa$ of $\Sigma$ restricting to $\eta'$ on $\Sigma_{\eta}$. The sum $\sum_{\sigma \in \Sigma} \sigma(\xi)\kappa(\sigma)^{-1}$ clearly lies in $\mathbb{Z}_{ur, \times}$ since its reduction modulo $p$ is the projection of the $\mathbb{F}_p[\Sigma]$-basis $\tilde{\xi}$ of $\mathcal{O}_F/(p) \otimes_{F_p} \mathbb{F}_p$ into the $\kappa$-eigenspace (up to the unit $|\Sigma| = f$), hence nonzero. So we find

$$\xi_{\eta'} \in \mathbb{Z}_{ur, \times}.$$  

We now analyze the second factor

$$x_{\eta} := \prod_{\tau \in \Sigma_{\eta}\backslash \Sigma} \sum_{\delta \in \Delta} \tau^{-1}\delta \cdot \eta(\delta)^{-1}$$

which is the product over the projections of $x$ into the $\eta^i$-eigenspaces for $i = 0, \ldots, f_{\eta} - 1$ (up to the unit $|\Delta| = e$). For $0 \leq j < e$ the $\eta_0^{-j}$-eigenspace of the inverse different is generated over $\mathcal{O}_F$ by $(\sqrt[6]{\mathcal{O}_F})^{-j}$ and since $x$ was a $\mathcal{O}_F[\Delta]$-basis of the inverse different its projection lies in $\mathcal{O}_F^\times \cdot (\sqrt[6]{\mathcal{O}_F})^{-j}$. So by Lemma 6 below we have

$$x_{\eta} \in \mathcal{O}_F^\times \cdot \prod_{i=0}^{f_{\eta}-1} (\sqrt[6]{\mathcal{O}_F})^{-e(p^i(-m_{\eta})/e)} \subset K$$

and hence

$$v_p(x_{\eta}) = -\sum_{i=0}^{f_{\eta}-1} \left( \frac{-m_{\eta}p^i}{e} \right) = -v_p(\tau(r_{\tilde{\eta}, \tilde{z}}, \psi_p)), \quad (15)$$

using Corollary 4 and the fact that $\tilde{\eta} = \eta_0^{-m_{\eta}}$. One checks that $\tau(r_{\tilde{\eta}, \tilde{z}}, \psi_p) \in \mathbb{Q}_{ur}^p(\xi_p)$ is an eigenvector for the character

$$\varphi = \eta_0^{-m_{\eta}(p^{f_{\eta} - 1})/(p-1)}.$$
of the group \( \text{Gal}(\mathbb{Q}_p^{\text{ur}}(\zeta_p) \cap K^{\text{ur}}/\mathbb{Q}_p) \). Also, since \( x_\eta \) is an eigenvector for \( \varrho^{-1} \), Equation (15) implies
\[ \tau(r_{\bar{\eta},\bar{z}}, \psi_p) \cdot x_\eta \in \mathbb{Z}_p^{\text{ur},\times}. \]
Combining this with (14) and (12) we find
\[ \epsilon(r_{\bar{\chi}}) \cdot \lbrack \text{per}(b) \rbrack_\chi = \tilde{\eta}(\text{rec}(p)) \tau(r_{\bar{\eta},\bar{z}}, \psi_p) \tilde{\eta}'(\sigma_{f_\eta}) \cdot x_\eta \cdot \xi_{\eta'} \epsilon \in \mathbb{Z}_p^{\text{ur},\times} \]
and hence
\[ (\epsilon(r_{\bar{\chi}}))_{\chi \in \hat{G}} \cdot \lbrack \text{per}(b) \rbrack \in \text{im}(K_1(\mathbb{Z}_p^{\text{ur}}[G])). \]

**Lemma 6.** We have \( \eta = \eta_0^{m_\eta} \), where \( \eta_0 \) is the character (10) associated to the element \( p_0 \) of valuation 1 and \( m_\eta \) was defined in (13).

**Proof.** It suffices to show that the composite map
\[ \omega' : \mu_{p^{f_{\eta}-1}} \subset F^\times \xrightarrow{\text{rec}} \mathcal{G}_F \rightarrow \text{Gal}(K/F) \xrightarrow{\eta_0^{m_\eta}} \mu_e \]
agrees with the \( (m_\eta(p^{f_{\eta}} - 1)/e) \)-th power map. By definition [Neukirch 1999, Theorem V.3.1] of the tame local Hilbert symbol and the fact that our map \( \text{rec} \) is the inverse of that used in [Neukirch 1999], we have
\[ \omega' (\zeta) = \left( \frac{\zeta^{-1} \cdot p_0^{m_\eta}}{F} \right), \]
which by [Neukirch 1999, Theorem V.3.4] equals
\[ \left( \frac{\zeta^{-1} \cdot p_0^{m_\eta}}{F} \right) = \left( -1 \right)^{\alpha \beta} \frac{p_0^\beta}{\zeta^{-\alpha}} \left( p^{f_{\eta}-1}/e \right) = \zeta^{m_\eta(p^{f_{\eta}-1})/e}, \]
where \( \alpha = v_p(p_0^{m_\eta}) = m_\eta \) and \( \beta = v_p(\zeta^{-1}) = 0. \)

Denote by \( \gamma \) a topological generator of
\[ \Gamma := \text{Gal}(\mathbb{Q}_p(\zeta_{p,\infty})/\mathbb{Q}_p) \]
and by
\[ \chi^{\text{cyclo}} : \text{Gal}(\mathbb{Q}_p(\zeta_{p,\infty})/\mathbb{Q}_p) \cong \mathbb{Z}_p^\times \]
the cyclotomic character. As in the proof of Proposition 5 choose \( b \) such that
\[ \mathbb{Z}_p[G] \cdot b = (\sqrt{p_0})^{-(e-1)} \mathcal{O}_K. \]
Denote by \( e_1 = \frac{1}{|\Sigma|} \sum_{g \in \Sigma} g \in \mathbb{Z}_p[\Sigma] \) the idempotent for the trivial character of \( \Sigma \).

**Proposition 7.** If \( p \nmid |G| \) then one can choose \( \beta \in H^1(K, \mathbb{Z}_p(1-r)) \) such that
\[ H^1(K, \mathbb{Z}_p(1-r)) = H^1(K, \mathbb{Z}_p(1-r))_{\text{tor}} \oplus \mathbb{Z}_p[G] \cdot \beta \]
and the local Tamagawa number conjecture (4) is equivalent to the identity

$$(r - 1)! \cdot (p^{(r-1)c(\chi)})_{\chi \in \hat{G}} \cdot [\text{per}(b)]^{-1} \cdot [\text{per}(\exp^*(\beta))] \cdot [C_\beta]^{-1} \cdot \left[ \frac{1 - p^{r-1}\sigma}{1 - p^{-r}\sigma^{-1}} \right] \equiv 1_F$$

in the group $K_1(\mathbb{Q}_p^{ur}[G]) \backslash \text{im} K_1(\mathbb{Z}_p^{ur}[G])$. The projection of this identity into the group $K_1(\mathbb{Q}_p^{ur}[\Sigma]) \backslash \text{im} K_1(\mathbb{Z}_p^{ur}[\Sigma])$ is

$$[(r - 1)!] \cdot [\text{per}(\exp^*(\beta))]_F \cdot \left[ \frac{\chi_{\text{cyclo}}(\gamma)^r - 1}{\chi_{\text{cyclo}}(\gamma)^r - 1} e_1 + 1 - e_1 \right] \cdot \left[ \frac{1 - p^{r-1}\sigma}{1 - p^{-r}\sigma^{-1}} \right] \equiv 1_F$$

and in the components of $K_1(\mathbb{Q}_p^{ur}[G]) \backslash \text{im} K_1(\mathbb{Z}_p^{ur}[G])$ indexed by $\chi = ([\eta], \eta')$ with

$$\eta|_{\text{Gal}(K/K \cap F(\xi_p))} \neq 1$$

this identity is equivalent to

$$(r - 1)! f_\eta \cdot p^{(r-1)f_\eta} \cdot [\text{per}(b)]_\chi^{-1} \cdot [\text{per}(\exp^*(\beta))]_\chi \in \mathbb{Z}_p^{ur, \infty}. \quad (16)$$

**Proof.** If $p \nmid |G|$ then the module $H^1(K, \mathbb{Z}_p(1-r))/\text{tor}$ is free over $\mathbb{Z}_p[G]$ since this is true for any lattice in a free rank-one $\mathbb{Q}_p[G]$-module. The first statement is then clear from (9) and **Proposition 5**.

Since

$$R \Gamma(K, \mathbb{Z}_p(1-r)) \otimes_{\mathbb{Z}_p[G]} \mathbb{Z}_p[\Sigma] \cong R \Gamma(F, \mathbb{Z}_p(1-r)),$$

the projection $[C_\beta]_F$ of $[C_\beta]$ into $K_1(\mathbb{Q}_p^{ur}[\Sigma]) \backslash \text{im} K_1(\mathbb{Z}_p^{ur}[\Sigma])$ is the class of the complex

$$H^1(F, \mathbb{Z}_p(1-r))_{\text{tor}}[-1] \oplus H^2(F, \mathbb{Z}_p(1-r))[-2]$$

and both modules have trivial $\Sigma$-action. Any finite cyclic $\mathbb{Z}_p[\Sigma]$-module $M$ with trivial $\Sigma$-action has a projective resolution

$$0 \to \mathbb{Z}_p[\Sigma] \xrightarrow{[M]e_1 + 1 - e_1} \mathbb{Z}_p[\Sigma] \to M \to 0$$

and the class of $M$ in $K_0(\mathbb{Z}_p[\Sigma], \mathbb{Q}_p)$ is represented by $[[M]e_1 + 1 - e_1]^{-1} \in K_1(\mathbb{Q}_p[\Sigma])$. Using Tate local duality we have

$$[C_\beta]_F = [H^1(F, \mathbb{Z}_p(1-r))_{\text{tor}}]^{-1} \cdot [H^2(F, \mathbb{Z}_p(1-r))]$$

$$= [H^0(F, \mathbb{Q}_p/\mathbb{Z}_p(1-r))]^{-1} \cdot [H^0(F, \mathbb{Q}_p/\mathbb{Z}_p(r))]$$

$$= \left[ (\chi_{\text{cyclo}}(\gamma)^{r-1} - 1)e_1 + 1 - e_1 \right] \cdot \left[ (\chi_{\text{cyclo}}(\gamma)^{r} - 1)e_1 + 1 - e_1 \right]^{-1}$$

$$= \left[ \frac{\chi_{\text{cyclo}}(\gamma)^{r-1} - 1}{\chi_{\text{cyclo}}(\gamma)^{r} - 1} e_1 + 1 - e_1 \right].$$
By Proposition 5 \([\text{per}(b)]_\chi\) is a \(p\)-adic unit if \(\eta = 1\), which gives the second statement. The third statement follows from the fact that \(\text{Gal}(K/K \cap F(\zeta_p))\) acts trivially on \(R\Gamma(K, \mathbb{Z}_p(1-r))\) which implies that \([C_\beta]_\chi = 1\) if the restriction of \(\eta\) to \(\text{Gal}(K/K \cap F(\zeta_p))\) is nontrivial. \(\square\)

4. The Cherbonnier–Colmez reciprocity law

Now that we have reformulated Equation (4) according to Proposition 7 we see that we must compute the image of \(\exp^*(\beta)\). In order to do this we will use an explicit reciprocity law of [Cherbonnier and Colmez 1999], which uses the theory of \((\phi,0\mathbb{K})\)-modules and the rings of periods of Fontaine. Rather than developing this machinery in full, we will give only the definitions and results needed to state the reciprocity in our case; the reader is invited to read [Cherbonnier and Colmez 1999] to see the theory and the reciprocity law developed in full generality.

4.1. Iwasawa theory. In this subsection and the next we recall results of [Cherbonnier and Colmez 1999] specialized to the representation \(V = \mathbb{Q}_p(1)\). For this discussion we temporarily suspend our assumption that \(p \nmid |G|\). So let \(K\) again be an arbitrary finite Galois extension of \(\mathbb{Q}_p\), define

\[
K_n = K(\zeta_{p^n}), \quad K_\infty = \bigcup_{n \in \mathbb{N}} K_n,
\]

and

\[
\Gamma_K := \text{Gal}(K_\infty/K), \quad \Lambda_K = \mathbb{Z}_p[[\text{Gal}(K_\infty/\mathbb{Q}_p)]]
\]

and

\[
H^m_{Iw}(K, \mathbb{Z}_p(1)) = \lim_{\leftarrow n} H^m(K_n, \mathbb{Z}_p(1)) \cong \lim_{\leftarrow n} H^m(K, \text{Ind}^{G_K}_{G_{K_n}} \mathbb{Z}_p(1)) \cong H^m(K, T),
\]

where the inverse limit is taken with respect to corestriction maps, the second isomorphism is Shapiro’s lemma and

\[
T := \lim_{\leftarrow n} \text{Ind}^{G_K}_{G_{K_n}} \mathbb{Z}_p(1) \cong \lim_{\leftarrow n} \mathbb{Z}_p[\text{Gal}(K_n/K)](1) \cong \mathbb{Z}_p[[\Gamma_K]](1)
\]

is a free rank-one \(\mathbb{Z}_p[[\Gamma_K]]\)-module with \(G_K\)-action given by \(\psi^{-1}\chi^{\text{cycl}}\), where

\[
\psi : G_K \to \Gamma_K \subseteq \mathbb{Z}_p[[\Gamma_K]]^\times
\]

is the tautological character (see the analogous discussion of (2)). From this it is easy to see that for any \(r \in \mathbb{Z}\) one has an exact sequence of \(G_K\)-modules

\[
0 \to T \xrightarrow{\gamma_K \cdot \chi^{\text{cycl}}(\gamma_K)^{-r-1}} T \to \mathbb{Z}_p(r) \to 0
\]
where \( \gamma_K \in \Gamma_K \) is a topological generator (our assumption that \( p \) is odd assures that \( \Gamma_K \) is procyclic for any \( K \)). It is clear from the definition that

\[
H^m_{Iw}(K, \mathbb{Z}_p(1)) \cong H^m_{Iw}(K_n, \mathbb{Z}_p(1))
\]

for any \( n \geq 0 \). So \( H^m_{Iw}(K, \mathbb{Z}_p(1)) \) only depends on the field \( K_{\infty} \), and it is naturally a \( \Lambda_K \)-module. Since our base field \( K \) was arbitrary an analogous sequence holds with \( K \) replaced by \( K_n \) and \( T \) by the corresponding \( G_{K_n} \)-module \( T_n \) so that \( T \cong \text{Ind}_{G_{K_n}}^{G_K} T_n \).

In view of (18) we obtain induced maps

\[
pr_{n,r} : H^1_{Iw}(K, \mathbb{Z}_p(1)) \to H^1(K_n, \mathbb{Z}_p(r))
\]

for any \( n \geq 0 \) and \( r \in \mathbb{Z} \).

**Lemma 8.** Set \( \gamma_n = \gamma_{K_n} \). If \( r \neq 1 \) then the map \( pr_{n,r} \) induces an isomorphism

\[
H^1_{Iw}(K, \mathbb{Z}_p(1))/(\gamma_n - \chi_c^{\text{cyclo}}(\gamma_n)^{1-r}) H^1_{Iw}(K, \mathbb{Z}_p(1)) \cong H^1(K_n, \mathbb{Z}_p(r)).
\]

**Proof.** The short exact sequence (17) over \( K_n \) induces a long exact sequence of cohomology groups

\[
0 \to H^0_{Iw}(K, \mathbb{Z}_p(1)) \xrightarrow{\gamma_n - \chi_c^{\text{cyclo}}(\gamma_n)^{1-r}} H^0_{Iw}(K, \mathbb{Z}_p(1)) \to H^0(K_n, \mathbb{Z}_p(r))
\]

\[
H^1_{Iw}(K, \mathbb{Z}_p(1)) \xrightarrow{\gamma_n - \chi_c^{\text{cyclo}}(\gamma_n)^{1-r}} H^1_{Iw}(K, \mathbb{Z}_p(1)) \to \text{pr}_{n,r} H^1(K_n, \mathbb{Z}_p(r))
\]

\[
H^2_{Iw}(K, \mathbb{Z}_p(1)) \xrightarrow{\gamma_n - \chi_c^{\text{cyclo}}(\gamma_n)^{1-r}} H^2_{Iw}(K, \mathbb{Z}_p(1)) \to H^2(K_n, \mathbb{Z}_p(r)) \to 0.
\]

By Tate local duality there is a canonical isomorphism of \( \text{Gal}(K_n/K) \)-modules

\[
H^2(K_n, \mathbb{Z}_p(1)) \cong \mathbb{Z}_p
\]

for each \( n \), and the corestriction map is the identity map on \( \mathbb{Z}_p \). Hence,

\[
H^2_{Iw}(K, \mathbb{Z}_p(1)) \cong \mathbb{Z}_p
\]

with trivial action of \( \Gamma_{K_n} \). This implies that for \( r \neq 1 \) multiplication by

\[
\gamma_n - \chi_c^{\text{cyclo}}(\gamma_n)^{1-r} = 1 - \chi_c^{\text{cyclo}}(\gamma_n)^{1-r}
\]

is injective on \( H^2_{Iw}(K, \mathbb{Z}_p(1)) \). Hence \( pr_{n,r} \) is surjective and we obtain the desired isomorphism. \( \square \)
4.2. The ring $A_K$ and the reciprocity law. The theory of $(\varphi, \Gamma_K)$-modules [Cherbonnier and Colmez 1999] involves a ring

$$A_K = \left( \mathcal{O}_{F'}[[\pi_K]][1/\pi_K] \right)^\wedge = \left\{ \sum_{n \in \mathbb{Z}} a_n \pi_K^n : a_n \in \mathcal{O}_{F'}, \lim_{n \to -\infty} a_n = 0 \right\},$$

where $\pi_K$ is (for now) a formal variable and $F' \supseteq F$ is the maximal unramified subfield of $K_\infty$. (The notation $(-)^\wedge$ means $\hat{-}$.) The ring $A_K$ carries an operator $\varphi$ extending the Frobenius on $\mathcal{O}_{F'}$ and an action of $\Gamma_K$ commuting with $\varphi$, which are somewhat hard to describe in terms of $\pi_K$. However, on the subring

$$A_{F'} = \left( \mathcal{O}_{F'}[[\pi]][1/\pi] \right)^\wedge \subseteq A_K$$

one has

$$\varphi(1 + \pi) = (1 + \pi)^p, \quad \gamma(1 + \pi) = (1 + \pi)^{\chi_{\text{cyclo}}(\gamma)}$$

for $\gamma \in \Gamma_K$.

The ring $A_K$ is a complete, discrete valuation ring with uniformizer $p$. We denote by $E_K \cong k((\pi_K))$ its residue field and by $B_K = A_K[1/p]$ its field of fractions. We see that $\varphi(B_K)$ is a subfield of $B_K$ (of degree $p$), and thus we can define

$$\psi = p^{-1} \varphi^{-1} \text{Tr}_{B_K/\varphi B_K}$$

and

$$\mathcal{N} = \varphi^{-1} N_{B_K/\varphi B_K}$$

as further operators on $B_K$. We observe that if $f \in B_K$, then

$$\psi(\varphi(f)) = f.$$

Thus $\psi$ is an additive left inverse of $\varphi$. We write $A_K^\psi = 1 \subset A_K$ for the set of elements fixed by the operator $\psi$. The $(\varphi, \Gamma_K)$-module associated to the representation $\mathbb{Z}_p(1)$ is $A_K(1)$ where the Tate twist refers to the $\Gamma_K$-action being twisted by the cyclotomic character.

By [Cherbonnier and Colmez 1999, III.2] the field $B_K$ is contained in a field $\tilde{B}$ on which $\varphi$ is bijective and $\tilde{B}$ contains a $G_K$-stable subring $\tilde{B}^{\dagger,n}$ consisting of elements $x$ for which $\varphi^{-n}(x)$ converges to an element in $B_{dR}$. So one has a $G_K$-equivariant ring homomorphism

$$\varphi^{-n} : \tilde{B}^{\dagger,n} \to B_{dR},$$

which again is rather inexplicit in general but is given by

$$\varphi^{-n}(\pi) = \zeta_p^n e^{i/p^n} - 1$$

on the element $\pi$.

The main result [Cherbonnier and Colmez 1999, théorème IV.2.1] specialized to the representation $V = \mathbb{Q}_p(1)$ can now be summarized as follows.
Theorem 9. Let $K/\mathbb{Q}_p$ be any finite Galois extension and

$$\Lambda_K := \mathbb{Z}_p[[\text{Gal}(K_\infty/\mathbb{Q}_p)]]$$

its Iwasawa algebra.

(a) There is an isomorphism of $\Lambda_K$-modules

$$\text{Exp}^*_{\mathbb{Z}_p^1} : H^1_{Iw}(K, \mathbb{Z}_p(1)) \cong A^{\psi=1}_K(1).$$

(b) There is $n_0 \in \mathbb{Z}$ such that for $n \geq n_0$ the following hold:

(b1) $A^{\psi=1}_K \subseteq \tilde{B}_{n}^\dagger$.

(b2) The $G_K$-equivariant map $\varphi^{-n} : A^{\psi=1}_K \to B_{dR}$ factors through

$$\varphi^{-n} : A^{\psi=1}_K \to K_n[[t]] \subseteq B_{dR}.$$

(b3) One has

$$p^{-n} \varphi^{-n}(\text{Exp}^*_{\mathbb{Z}_p}(u)) = \sum_{r=1}^{\infty} \exp_{\mathbb{Q}_p(r)}(\text{pr}_{1-r}(u)) \cdot t^{r-1}$$

for any $u \in H^1_{Iw}(K, \mathbb{Z}_p(1))$.

Theorem 9 contains all the information we shall need when analyzing the case of tamely ramified $K$ in Section 6 below. However, the paper [Cherbonnier and Colmez 1999] contains further information on the map $\text{Exp}^*_{\mathbb{Z}_p}$, which we summarize in the next proposition. We shall only need this proposition when reproving the unramified case of the local Tamagawa number in Section 5 below. First recall from [Cherbonnier and Colmez 1999, p. 257] that the ring $B_K$ carries a derivation

$$\nabla : B_K \to B_K,$$

uniquely specified by its value on $\pi$:

$$\nabla(\pi) = 1 + \pi.$$

We set

$$\nabla \log(x) = \frac{\nabla(x)}{x}$$

and denote by

$$\widehat{M} := \lim_{\rightarrow} M/p^n M$$

the $p$-adic completion of an abelian group $M$. 
**Proposition 10.** There is a commutative diagram of \( \Lambda_K \)-modules, where the maps labeled by \( \cong \) are isomorphism.

\[
\begin{array}{ccc}
A(K_\infty) := \lim_{\longleftarrow m,n} K_n^\times / (K_n^\times)^{p^m} & \cong & (E_K^\times)^\wedge \\
\uparrow & \cong & \uparrow \\
U := \lim_{\longleftarrow m,n} O_{K_n}^\times / (O_{K_n}^\times)^{p^m} & \cong & 1 + \overline{\pi}_K k[[\overline{\pi}_K]]
\end{array}
\]

**Proof.** The isomorphism \( \delta \) arises from Kummer theory. The theory of the field of norms gives an isomorphism of multiplicative monoids [Cherbonnier and Colmez 1999, proposition I.1.1]

\[\lim_n O_{K_n} \cong k[[\overline{\pi}_K]],\]

which induces our isomorphism \( \iota_K |_U \) after restricting to units and passing to \( p \)-adic completions and our isomorphism \( \iota_K \) by taking the field of fractions and passing to \( p \)-adic completions of its units.

By [Cherbonnier and Colmez 1999, corollaire V.1.2] (see also [Daigle 2014, 3.2.1] for more details), the reduction-mod-\( p \)-map \( (A_K^{N=1}) ^\wedge \rightarrow (E_K^\times)^\wedge \) is an isomorphism.

By [Cherbonnier and Colmez 1999, proposition V.3.2(iii)] the map \( \nabla \log \) makes the upper triangle in our diagram commute. Since all other maps in this triangle are isomorphisms, the map \( \nabla \log \) is an isomorphism as well. \( \square \)

**4.3. Specialization to the tamely ramified case.** We now resume our assumption that \( p \) does not divide the degree of \( [K : Q_p] \) together with (most of) the notation from Section 3. In addition we assume that

\[\zeta_p \in K,\]

which implies that \( K_{\infty}/K \) is totally ramified and hence that \( F = F' \) is the maximal unramified subfield of \( K_{\infty} \). The theory of fields of norms [Cherbonnier and Colmez 1999, remarque I.1.2] shows that \( E_K \) is a Galois extension of \( E_F \) of degree

\[e := [K_{\infty} : F_{\infty}] = [K : F(\zeta_p)]\]

with group

\[\text{Gal}(E_K/E_F) \cong \text{Gal}(K_{\infty}/F_{\infty}) \cong \text{Gal}(K/F(\zeta_p)).\]

Note that with this notation the ramification degree of \( K/\mathbb{Q}_p \) is \( e(p - 1) \) whereas it was denoted by \( e \) in Section 3. The element \( p_0 \) of Section 3 we choose to be \(-p\),
i.e., we assume that

\[ K = F(\sqrt[p-1]{-p}). \]

An easy computation shows that \((\zeta_p - 1)^{p-1} = -p \cdot u\) with \(u \in 1 + (\zeta_p - 1)\mathbb{Z}_p[\zeta_p]\)
and hence we can choose the root \(\sqrt[p-1]{-p}\) such that

\[ \zeta_p - 1 = (\sqrt[p-1]{-p}) \cdot u' \]

with \(u' \in 1 + (\zeta_p - 1)\mathbb{Z}_p[\zeta_p]\). By Kummer theory we then also have

\[ K = F(\sqrt[p-1]{\zeta_p - 1}) \]

and \(B_K = B_F(\sqrt{\pi})\). Any choice of \(\pi_K = \sqrt{\pi}\) fixes a choice of

\[ \sqrt[p-1]{\zeta_p - 1} = \varphi^{-1}(\pi_K)|_{t=0} \]

and of

\[ e^{(p-1)/p} = \sqrt[p-1]{\zeta_p - 1} \cdot (u')^{-1/e}. \]

We have

\[ G \cong \Sigma \ltimes \Delta \]

with \(\Sigma\) cyclic of order \(f\) and \(\Delta\) cyclic of order \(e(p-1)\) and

\[ \Lambda_K \cong \mathbb{Z}_p[[G \times \Gamma_K]] \cong \mathbb{Z}_p[\Sigma \ltimes \Delta][[\gamma_1 - 1]], \]

where \(\gamma_1 = \gamma^{p-1}\) is a topological generator of \(\Gamma_K\).

**Proposition 11.** There is an isomorphism of \(\Lambda_K\)-modules

\[ H^1_{Iw}(K, \mathbb{Z}_p(1)) \cong \Lambda_K \cdot \beta_{Iw} \oplus \mathbb{Z}_p(1). \]

**Proof.** In view of the Kummer theory isomorphism

\[ \delta : A(K_\infty) \cong H^1_{Iw}(K, \mathbb{Z}_p(1)) \]

it suffices to quote the structure theorem for the \(\Lambda_K\)-module \(A(K_\infty)\) given in
[Neukirch et al. 2000, Theorem 11.2.3](#) (where \(k = \mathbb{Q}_p\) and our group \(\Sigma \ltimes \Delta\) is the group \(\Delta\) of [loc. cit.]).

**Corollary 12.** There is an isomorphism of \(\mathbb{Z}_p[G]\)-modules

\[ H^1(K, \mathbb{Z}_p(1-r)) \cong \mathbb{Z}_p[G] \cdot \beta \oplus H^1(K, \mathbb{Z}_p(1-r))_{\text{tor}}, \]

where \(\beta = \text{pr}_{0,1-r}(\beta_{Iw}) = \text{pr}_{1,1-r}(\beta_{Iw}).\)
Proof. This is clear from Proposition 11 and Lemma 8 (with \( r \) replaced by \( 1 - r \)) in view of the isomorphisms

\[
\mathbb{Z}_p[G] \cong \Lambda_K / (\gamma_1 - \chi^{\text{cyclo}}(\gamma_1)^r) \Lambda_K
\]

and

\[
\mathbb{Z}_p(1) / (\gamma_1 - \chi^{\text{cyclo}}(\gamma_1)) \mathbb{Z}_p(1) = \mathbb{Z}_p / (\chi^{\text{cyclo}}(\gamma_1) - \chi^{\text{cyclo}}(\gamma_1)^r) \mathbb{Z}_p \\
\cong H^0(K, \mathbb{Q}_p / \mathbb{Z}_p(1 - r))) \\
\cong H^1(K, \mathbb{Z}_p(1 - r))_{\text{tor}}.
\]

If we choose the element \( \beta \) of Corollary 12 to verify the identity in Proposition 7 it remains to get an explicit hold on some \( \mathbb{Z}_p \)-basis \( \beta_{Iw} \), or rather of its image

\[
\alpha = \text{Exp}^*_{\mathbb{Z}_p}(\beta_{Iw}) \in \Lambda_K^{\psi=1}(1).
\] (22)

Since \( \alpha \) is a (infinite) Laurent series in \( \pi_K \) it will be amenable to somewhat explicit analysis. In the unramified components of Proposition 7 (\( \eta = 1 \)) we can compute \( \alpha \) in terms of the well-known Perrin-Riou basis (see Proposition 24 below) which is a main ingredient in all known proofs of the unramified case of the local Tamagawa number conjecture. In the other components (\( \eta \neq 1 \)) we shall simply use Nakayama’s lemma to analyze \( \alpha \) as much as we can in Section 6.

In order to compute \( \text{exp}^*_{\mathbb{Q}_p}(\beta) \) we also need to be able to apply Theorem 9 for \( n = 1 \).

\textbf{Proposition 13.} Part (b) of Theorem 9 holds with \( n_0 = 1 \).

\textbf{Proof.} It will follow from an explicit analysis of elements in \( \Lambda_K^{\psi=1} \) in Corollary 37 below that \( \varphi^{-1}(a) \) converges for \( a \in \Lambda_K^{\psi=1} \), which shows (b1). Since \( \pi_K^e = \pi \) and \( \varphi^{-n}(\pi) = \zeta_p^n e^{t/\ell^{n}} - 1 \) it is also clear that the values of \( \varphi^{-n} \) on \( \Lambda_K \), if convergent, lie in \( F(\sqrt[n]{\zeta_p^n - 1})[[t]] = K_n[[t]] \). This shows (b2). By [Cherbonnier and Colmez 1999, théorème IV.2.1] the right-hand side of (b3) is given by \( T_n \varphi^{-m}(\text{Exp}^*_{\mathbb{Z}_p}(u)) \) for \( m \geq n \) large enough (see the next section for the definition of \( T_n \)). The statement in (b3) then follows from Corollary 17 below. \( \square \)

\textbf{4.4. Some power series computations.} The purpose of this section is simply to record some computations justifying Theorem 9(b3) for \( n \geq 1 \). Another aim is to write the coefficients of the right-hand side of Theorem 9(b3) in terms of the derivation \( \nabla \) applied to the left-hand side. First we have

\textbf{Lemma 14 [Cherbonnier and Colmez 1999, lemme III.2.3].} Suppose \( \varphi^{-n}f \) and \( \varphi^{-n}(\nabla f) \) both converge in \( B_{\text{der}} \). Then

\[
\varphi^{-n}(\nabla f) = p^n \frac{d}{dt}(\varphi^{-n}(f)).
\]
Proof. It’s enough to check that $\varphi^{-n} \circ \nabla$ and $p^n \frac{d}{dt} \varphi^{-n}$ both agree on $1 + \pi$, since they are both derivations. We see that

\[
\varphi^{-n} \nabla (1 + \pi) = \varphi^{-n} (1 + \pi) = \zeta p^ne^{t/p^n}
\]
\[
p^n \frac{d}{dt} \varphi^{-n} (1 + \pi) = p^n \frac{d}{dt} \zeta p^ne^{t/p^n} = \zeta p^ne^{t/p^n}.
\]

□

The next Lemma shows that $\nabla$ is compatible with other operators that we have introduced. The ring $B$ is defined as in [Cherbonnier and Colmez 1999].

**Lemma 15.** Let $f \in B_K$. Then we have

(a) $\nabla \chi = \chi \cdot \gamma \nabla f$,
(b) $\nabla \varphi f = p \cdot \varphi \nabla f$,
(c) $\nabla \text{Tr}_{B/\varphi B} f = \text{Tr}_{B/\varphi B} \nabla f$,
(d) $\nabla \psi f = p^{-1} \cdot \psi \nabla f$.

**Proof.** This is a straightforward computation. For example, to see (c) note that $(1 + \pi)^i$, $i = 0, \ldots, p - 1$ is a $\varphi B$-basis of $B$ and

\[
\text{Tr}_{B/\varphi B}(x) = \text{Tr}_{B/\varphi B} \left( \sum_{i=0}^{p-1} \varphi x_i \cdot (1 + \pi)^i \right) = p \cdot \varphi x_0.
\]

Hence

\[
\text{Tr}_{B/\varphi B}(\nabla x) = \text{Tr}_{B/\varphi B} \left( \sum_{i=0}^{p-1} \nabla \varphi x_i \cdot (1 + \pi)^i + \varphi x_i \cdot i \cdot (1 + \pi)^i \right)
\]
\[
= \text{Tr}_{B/\varphi B} \left( \sum_{i=0}^{p-1} \varphi (p \nabla x_i + x_i \cdot i) \cdot (1 + \pi)^i \right)
\]
\[
= p^2 \varphi \nabla x_0 = \nabla (p \cdot \varphi x_0) = \nabla \text{Tr}_{B/\varphi B}(x).
\]

See [Daigle 2014, Lemma 3.1.3] for more details. □

Recall the normalized trace maps

\[
T_n : K_\infty \rightarrow K_n
\]

from [Cherbonnier and Colmez 1999, p. 259] which are given by

\[
T_n(x) = p^{-m} \text{Tr}_{K_m/K_n} x
\]

for any $m \geq n$ such that $x \in K_m$, and extend to a map

\[
T_n : K_\infty[[t]] \rightarrow K_n[[t]]
\]

by linearity. By [Cherbonnier and Colmez 1999, théorème IV.2.1] the right-hand side of Theorem 9(b3) is given by $T_n \varphi^{-m}(f)$ for $f = \text{Exp}_p^*(u) \in A_K^{\psi=1}$ and $m \geq n$.
large enough. In order to get access to individual Taylor coefficients of the right-hand side we wish to compute \( \frac{d^{r-1}}{dt^{r-1}} T_n \varphi^{-m}(f) \), but from Lemmas 14 and 15 we see that

\[
\frac{d^{r-1}}{dt^{r-1}} T_n \varphi^{-m} = p^{-m(r-1)} T_n \varphi^{-m} \nabla^{r-1}
\]

and thus we can study the map \( T_n \varphi^{-m} \) on \( \nabla^{r-1} A_K^{\psi=1} \). But since \( \psi \nabla x = p \nabla \psi x \), we see that \( \nabla^{r-1} A_K^{\psi=1} \subseteq A_K^{\psi=p^{r-1}} \), and so we wish to study \( T_n \varphi^{-m} \) on \( A_K^{\psi=p^{r-1}} \).

**Lemma 16.** Let \( P \in A_K^{\psi=p^{-1}} \) be such that

\[
(\varphi^{-n} P)(0) := \varphi^{-n} P|_{t=0}
\]

converges and assume \( m \geq n \). Then if \( n \geq 1 \) we have

\[
(T_n \varphi^{-m} P)(0) = p^{(r-1)m-rn}(\varphi^{-n} P)(0). \tag{23}
\]

and if \( n = 0 \) we have

\[
(T_0 \varphi^{-m} P)(0) = p^{(r-1)m}(1 - p^{-r} \sigma^{-1})(\varphi^{-n} P)(0). \tag{24}
\]

**Proof.** Since \( P \in A_K^{\psi=p^{-1}} \), we know that \( \psi(P) = p^{-1} P \) and thus that

\[
p^{-r} \text{Tr}_{B/\varphi B}(P) = \varphi(P).
\]

Recall that we can choose \( \pi_K \) such that \( \pi_K^e = \pi \). Then \( \{((1+\pi)\xi-1)^{1/e} : \xi \in \mu_p\} \) is the set of conjugates of \( \pi_K \) over \( \varphi(B) \) in an algebraic closure of \( B \), so this gives

\[
p^{-r} \sum_{\xi \in \mu_p} P((1+\pi)\xi-1)^{1/e} = P^\sigma((1+\pi)^p-1)^{1/e}).
\]

Whenever \( \varphi^{-(l+1)} P \) converges for some \( l \in \mathbb{N} \), the operator \( \varphi^{-(l+1)} P|_{t=0} \) corresponds to setting \( \pi = \xi_{p^{l+1}} - 1 \) and applying \( \sigma^{-(l+1)} \) to each coefficient. We get

\[
p^{-r} \sum_{\xi \in \mu_p} P^{\sigma^{-(l+1)}}((\xi \cdot \xi_{p^{l+1}} - 1)^{1/e}) = P^{\sigma^{-l}}((\xi_{p^{l}} - 1)^{1/e})). \tag{25}
\]

If \( l \geq 1 \), this simplifies to

\[
p^{-r} \text{Tr}_{K_{l+1}/K_l} P^{\sigma^{-(l+1)}}((\xi_{p^{l+1}} - 1)^{1/e}) = P^{\sigma^{-l}}((\xi_{p^{l}} - 1)^{1/e}),
\]

and by induction, we see that for any \( 1 \leq n < m \),

\[
p^{m-r(m-n)} T_n P^{\sigma^{-m}}((\xi_{p^{n}} - 1)^{1/e}) = P^{\sigma^{-n}}((\xi_{p^{n}} - 1)^{1/e}). \tag{26}
\]
Since $P^\sigma((\xi_p^m - 1)^{1/e}) = (\varphi^{-m} P)(0)$, this proves Equation (23). If $l = 0$ then Equation (25) becomes

$$p^{-r} \sum_{\xi \in \mu_p} P^{\sigma^{-1}}((\xi \cdot \xi_p - 1)^{1/e}) = (\varphi^{-0} P)(0).$$

The left-hand side is now equal to

$$p^{-r} P^{\sigma^{-1}}(0) + p^{-r} \text{Tr}_{K_1/K_0} P^{\sigma^{-1}}((\xi - 1)^{1/e})$$

and we have

$$p^{-r} \text{Tr}_{K_1/K_0}(P^{\sigma^{-1}}((\xi - 1)^{1/e})) = (1 - p^{-r} \sigma^{-1})(\varphi^{-0} P)(0).$$

By induction we get

$$p^{m-rm} T_0 P^{\sigma^{-m}}((\xi_p^m - 1)^{1/e}) = (1 - p^{-r} \sigma^{-1})(\varphi^{-0} P)(0),$$

which proves Equation (24).

Corollary 17. If $P \in A^\psi_{K=1}$ is such that $\varphi^{-n} P$ converges and $m \geq n$, then we have

$$T_n \varphi^{-m} P = p^{-n} \varphi^{-n} P$$

if $n \geq 1$, and

$$T_0 \varphi^{-m} P = (1 - p^{-1} \sigma^{-1}) \varphi^{-0} P$$

if $n = 0$.

Proof. This follows by combining Lemma 16 for all $r$. □

5. The unramified case

In this section we reprove the local Tamagawa number conjecture (4) in the case where $K = F$ is unramified over $\mathbb{Q}_p$. This was first proven in [Bloch and Kato 1990] and other proofs can be found in [Perrin-Riou 1994; Benois and Berger 2008]. The proofs differ in the kind of “reciprocity law” which they employ but all proofs, including ours, use the “Perrin-Riou basis,” i.e., the $\Lambda_F$-basis in Proposition 24 below.

5.1. An extension of Proposition 10 in the unramified case. In this section we use results of Perrin-Riou [1990] to extend the diagram in Proposition 10 to the
diagram in Corollary 21 below. Define

$$P_F := \left\{ \sum_{n \geq 0} a_n \pi^n \in F[[\pi]] : na_n \in O_F \right\},$$

$$\overline{P}_F := P_F / pO_F[[\pi]],$$

$$\overline{P}_{F,\log} := \{ f \in \overline{P}_F : (p - \varphi)(f) = 0 \},$$

$$P_{F,\log} := \{ f \in P_F : \hat{f} \in \overline{P}_{F,\log} \} = \{ f \in P_F : (p - \varphi)(f) \in pO_F[[\pi]] \},$$

$$O_F[[\pi]]_{\log} := \{ f \in O_F[[\pi]] \times : f \mod pO_F[[\pi]] \in 1 + \pi k[[\pi]] \}
= 1 + (\pi, p).$$

Note that $P_F$ is the space of power series in $F$ whose derivative with respect to $\pi$ lies in $O_F[[\pi]]$. Observe that the map $d\log$ is given by an integral power series, and therefore $\log O_F[[\pi]]_{\log} \subseteq P_F$ where the logarithm map

$$\log(1 + x) = \sum_{n \geq 1} (-1)^{n-1} \frac{x^n}{n}$$

is given by the usual power series. Since $\varphi$ reduces modulo $p$ to the Frobenius, i.e., to the $p$-th power map, the logarithm series in fact induces a map

$$\log : O_F[[\pi]]_{\log} \rightarrow P_{F,\log}.$$

We wish to show that this map is an isomorphism, and to do this we first recall a couple of lemmas from [Perrin-Riou 1990].

**Lemma 18** [Perrin-Riou 1990, lemme 2.1]. Let

$$f \in 1 + \pi k[[\pi]] = \widehat{G}_m(k[[\pi]])$$

and let $\hat{f}$ be any lift of $f$ to $O_F[[\pi]]_{\log}$. Then

$$\log(\hat{f}) \mod pO_F[[\pi]] \in \overline{P}_{F,\log}$$

does not depend on the choice of $\hat{f}$, and the map $f \mapsto \log(\hat{f}) \mod pO_F[[\pi]]$ is an isomorphism $\log_k : 1 + \pi k[[\pi]] \xrightarrow{\sim} \overline{P}_{F,\log}$.

**Lemma 19** [Perrin-Riou 1990, lemme 2.2]. Let $f \in P_{F,\log}$. Then the sequence $p^m \psi^m(f)$ converges to a limit $f^\infty \in P_{F,\log}$, and we have

1. $f^\infty \equiv f \mod pO_F[[\pi]],$
2. $\psi(f^\infty) = p^{-1} f^\infty,$
3. $(1 - p^{-1} \varphi) f^\infty \in O_F[[\pi]],$
4. $f^\infty = 0 \text{ if } f \in O_F[[\pi]],$
5. $f^\infty = g^\infty \text{ if } f \equiv g \mod pO_F[[\pi]].$
Corollary 20.  (1) The map $\log : \mathcal{O}_F[[\pi]]_{\log} \to \mathcal{P}_{F,\log}$ is an isomorphism.

(2) One has a commutative diagram of isomorphisms

$$
\begin{array}{ccc}
\mathcal{O}_F[[\pi]]_{\log} & \xrightarrow{\log} & \mathcal{P}_{\psi=p^{-1}} \\
\cong \mod p & \cong \mod p & \\
1 + \pi k[[\pi]] & \xrightarrow{\log_k} & \overline{\mathcal{P}}_{F,\log}
\end{array}
$$

Proof. To see the first part, note that we have a commutative diagram

$$
\begin{array}{ccc}
1 & \to & 0 \\
\downarrow & & \downarrow \\
1 + p\mathcal{O}_F[[\pi]] & \xrightarrow{\log} & p\mathcal{O}_F[[\pi]] \\
\downarrow & & \downarrow \\
\mathcal{O}_F[[\pi]]_{\log} & \xrightarrow{\log} & \mathcal{P}_{F,\log} \\
\downarrow & & \downarrow \\
1 + \pi k[[\pi]] & \xrightarrow{\log_k} & \overline{\mathcal{P}}_{F,\log} \\
\downarrow & & \downarrow \\
1 & \to & 0
\end{array}
$$

and that the logarithm map on $1 + p\mathcal{O}_F[[\pi]]$ is an isomorphism since its inverse is given by the exponential series. By the five lemma, the middle arrow is an isomorphism. For the second part, it suffices to note that Lemma 19 shows that any element in $\overline{\mathcal{P}}_{F,\log}$ has a unique lift in $\mathcal{P}_{\psi=p^{-1}}$ and that $\log \mathcal{N}(x) = p\psi \log(x)$. □

Corollary 21. For $K = F$ the commutative diagram from Proposition 10 extends to a commutative diagram of $\Lambda_F$-modules:

$$
A(F_{\infty}) = \lim_{m,n} \frac{F_n}{(F_n)^{p^m}} \xrightarrow{\cong} (E_F^\times)^\wedge \xrightarrow{\cong} (A_F^{\psi=1})^\wedge \xrightarrow{\vee \log} A_F^{\psi=1}
$$

$$
U = \lim_{m,n} \mathcal{O}^\times_F/((\mathcal{O}^\times_F)^{p^m}) \xrightarrow{\cong} 1 + \pi k[[\pi]] \xrightarrow{\cong} \mathcal{O}_F[[\pi]]_{\log}^{\psi=p^{-1}} \xrightarrow{\log} \mathcal{P}_{F,\log}
$$

Proof. This is immediate from Corollary 20(2). □
This diagram allows us to determine the exact relationship between $P^{\psi=p^{-1}}_F$ and $A^{\psi=1}_F(1)$ since the relationship between $A(F_{\infty})$ and $U$ is quite transparent. There is an exact sequence of $\Lambda_F$-modules

$$0 \to U \to A(F_{\infty}) \xrightarrow{v} \mathbb{Z}_p \to 0,$$

where $v$ is the valuation map and $\mathbb{Z}_p$ carries the trivial $\Sigma \times \Gamma$-action. By [Neukirch et al. 2000, Theorem 11.2.3], already used in the proof of Proposition 11, there is an isomorphism

$$A(F_{\infty}) \cong \Lambda_F \oplus \mathbb{Z}_p(1) \quad (27)$$

and the torsion submodule $\mathbb{Z}_p(1)$ is clearly contained in $U$. Hence we obtain an exact sequence

$$0 \to U_{tf} \to A(F_{\infty})_{tf} \xrightarrow{v} \mathbb{Z}_p \to 0,$$

where $M_{tf} := M/M_{tors}$. The module $A(F_{\infty})_{tf}$ is free of rank one and since the $(\Sigma \times \Gamma)$-action on $\mathbb{Z}_p$ is trivial we find

$$U_{tf} = I \cdot A(F_{\infty})_{tf},$$

where

$$I := (\sigma - 1, \gamma - 1) \subseteq \Lambda_F$$

is the augmentation ideal.

**Lemma 22.** The augmentation ideal $I$ is principal, generated by the element

$$(1 - e_1) + (\gamma - 1)e_1,$$

where $e_1 \in \mathbb{Z}_p[\Sigma]$ is the idempotent for the trivial character of $\Sigma$.

**Proof.** This hinges on our assumption that $p$ does not divide the order of $\Sigma$, which implies that $e_1$ has coefficients in $\mathbb{Z}_p$. Using $e_1^2 = e_1$ we then find immediately

$$\sigma - 1 = (\sigma - 1)(1 - e_1) = (\sigma - 1)(1 - e_1) \cdot [(1 - e_1) + (\gamma - 1)e_1],$$

$$\gamma - 1 = ((\gamma - 1)(1 - e_1) + e_1) \cdot [(1 - e_1) + (\gamma - 1)e_1].$$

$\square$

**Lemma 23.** There are elements $\alpha \in A^{\psi=1}_F(1)$, $\tilde{\alpha} \in P^{\psi=p^{-1}}_{F, \log}$ such that

1. $A^{\psi=1}_F(1) = \Lambda_F \cdot \alpha \oplus \mathbb{Z}_p(1) \cdot 1$,
2. $P^{\psi=p^{-1}}_{F, \log} = \Lambda_F \cdot \tilde{\alpha} \oplus \mathbb{Z}_p \cdot \log(1 + \pi)$,
3. $\nabla \tilde{\alpha} = ((1 - e_1) + (\gamma - 1)e_1) \cdot \alpha$. 


Proof. Part (1) follows from (27) and Corollary 21. For part (2) one checks easily that $\mathbb{Z}_p \cdot \log(1 + \pi)$ is the torsion submodule of $P_{F, \log}^{\psi=p^{-1}}$ and that $(P_{F, \log}^{\psi=p^{-1}})_{tF}$ is free of rank one over $\Lambda_F$, since it is isomorphic under $\nabla$ to the free module 

$$I \cdot \alpha = \Lambda_F \cdot ((1 - e_1) + (\gamma - 1)e_1) \cdot \alpha$$

by Lemma 22. Note that we view $\alpha$ here as an element of $A_F(1)$, i.e., the action of $\gamma$ is $\chi_{\text{cyclo}}(\gamma)$ times the standard action (20) of $\gamma$ on $A_F$. Setting 

$$\tilde{\alpha} := \nabla^{-1}((1 - e_1) + (\gamma - 1)e_1) \cdot \alpha$$

we obtain (3). $\square$

5.2. The Coleman exact sequence and the Perrin-Riou basis. Lemma 23 tells us that $(P_{\psi=p^{-1}}^{\psi=p^{-1}})_{tF}$ is generated over $\mathbb{Z}_p$ by a single element $\tilde{\alpha}$, but not what this $\tilde{\alpha}$ is. By studying one more space, $O_F[[\pi]]^{\psi=0}$, we are able to describe $\tilde{\alpha}$ and hence $\alpha$.

Proposition 24. (1) There is an exact sequence of $\Lambda_F$-modules

$$0 \to \mathbb{Z}_p \cdot \log(1 + \pi) \to P_{F, \log}^{\psi=p^{-1}} \to O_F[[\pi]]^{\psi=0} \to \mathbb{Z}_p(1) \to 0. \tag{28}$$

(2) $O_F[[\pi]]^{\psi=0}$ is a free $\Lambda_F$-module of rank one generated by $\xi(1 + \pi)$, where $\xi \in O_F$ is a basis of $O_F$ over $\mathbb{Z}_p[\Sigma]$.

Proof. Part (1) is Theorem 2.3 in [Perrin-Riou 1990] and goes back to Coleman [1979]. See also [Daigle 2014, Proposition 4.1.10]. Part (2) is Lemma 1.5 in [Perrin-Riou 1990]. $\square$

Corollary 25. The bases $\alpha$ and $\tilde{\alpha}$ in Lemma 23 can be chosen such that

$$(1 - \varphi/p) \cdot \tilde{\alpha} = ((1 - e_1) + (\gamma - \chi_{\text{cyclo}}(\gamma))e_1) \cdot \xi(1 + \pi). \tag{29}$$

Proof. The cokernel of $(1 - \varphi/p)$ in (28) is isomorphic to

$$\mathbb{Z}_p(1) \cong \Lambda_F/(\sigma - 1, \gamma - \chi_{\text{cyclo}}(\gamma))$$

so the image of $(1 - \varphi/p)$ must be $(\sigma - 1, \gamma - \chi_{\text{cyclo}}(\gamma)) \cdot \xi(1 + \pi)$. As in Lemma 22 we can show that this ideal is principal, and is generated by

$$(1 - e_1) + (\gamma - \chi_{\text{cyclo}}(\gamma))e_1. \square$$

5.3. Proof of the conjecture for unramified fields. We now have the tools we need to explicitly compute $\exp^*_{Q_p(1)}(H^1(F, \mathbb{Z}_p(1 - r)))$ and prove the equality of Proposition 7 for $K = F$ (i.e., $e = 1$). By Lemma 8 we can take

$$\beta := \text{pr}_{0,1-r}(\beta_{1w}),$$
where $\beta_{lw}$ satisfies

$$\alpha = \text{Exp}^*_Z(\beta_{lw}),$$

$$\nabla \tilde{\alpha} = ((1 - e_1) + (\gamma - 1)e_1) \cdot \alpha,$$

$$1 - \varphi/p \cdot \tilde{\alpha} = ((1 - e_1) + (\gamma - \chi_{cyclo}(\gamma))e_1) \cdot \xi(1 + \pi),$$

using (22), Lemma 23(3) and (29). We cannot immediately apply Theorem 9 to $n = 0$, but going back to [Cherbonnier and Colmez 1999, théorème IV.2.1] we have

$$\sum_{r=1}^{\infty} \text{exp}^*_{Q_p(r)}(pr_{0,1-r}(u)) \cdot t^{r-1} = T_0 \varphi^{-m} \text{Exp}^*_Z(u).$$

Applying this to

$$u = ((1 - e_1) + (\gamma - 1)e_1) \cdot \beta_{lw}$$

assures that

$$\text{Exp}^*_Z(u) = \nabla \tilde{\alpha} \in \mathcal{O}_F[[\pi]]$$

and therefore

$$\varphi^{-0} P := \varphi^{-0} \nabla^{-1} \text{Exp}^*_Z(u) = \varphi^{-0} \nabla^{-1} \tilde{\alpha}$$

converges in $B_{\text{dR}}$ for any $r \geq 1$. Lemma 16 then implies

$$\text{exp}^*_{Q_p(r)}(pr_{0,1-r}(u)) = \frac{1}{(r-1)!} \left( \frac{d}{dt} \right)^{r-1} T_0 \varphi^{-m} \text{Exp}^*_Z(u) \bigg|_{t=0}$$

$$= \frac{1}{(r-1)!} T_0 p^{-(r-1)m} \varphi^{-m} \nabla^{r-1} \text{Exp}^*_Z(u) \bigg|_{t=0}$$

$$= \frac{1}{(r-1)!} (1 - p^{-r} \sigma^{-1}) \varphi^{-0} \nabla^{-1} \tilde{\alpha} \bigg|_{t=0}$$

$$= \frac{1}{(r-1)!} (1 - p^{-r} \sigma^{-1}) \nabla^{-1} \tilde{\alpha} \bigg|_{t=0}. $$

Applying $\nabla^{r}$ to (29) and using Lemma 15 we have

$$(1 - p^{r-1} \varphi) \cdot \nabla^{r} \tilde{\alpha} = \left( (1 - e_1) + (\chi_{cyclo}(\gamma)^{r} - \chi_{cyclo}(\gamma))e_1 \right) \cdot \nabla^{r} \xi(1 + \pi)$$

$$= \left( (1 - e_1) + (\chi_{cyclo}(\gamma)^{r} - \chi_{cyclo}(\gamma))e_1 \right) \cdot \xi(1 + \pi)$$

and so we find

$$\text{exp}^*_{Q_p(r)}(pr_{0,1-r}(u))$$

$$= \frac{1}{(r-1)!} \cdot \frac{1 - p^{-r} \sigma^{-1}}{1 - p^{-r} \sigma} \cdot \left( (1 - e_1) + (\chi_{cyclo}(\gamma)^{r} - \chi_{cyclo}(\gamma))e_1 \right) : \xi.$$
By Lemma 8 the action of $\gamma \in \Lambda_F$ on $H^1(F, \mathbb{Z}_p(1 - r))$ is via the character $\chi_{\text{cyclo}}(\gamma)^r$. Hence, for our choice (31) of $u$, we have

$$\text{pr}_{0,1-r}(u) = ((1 - e_1) + (\chi_{\text{cyclo}}(\gamma)^r - 1)e_1) \cdot \text{pr}_{0,1-r}(\beta_{Iw})$$

and we can finally compute

$$\exp_{Q_p(r)}^\ast(\beta) = \frac{1}{(r - 1)!} \cdot \frac{1 - p^{-r} \sigma^{-1}}{1 - p^{r-1} \sigma} \cdot \frac{(1 - e_1) + (\chi_{\text{cyclo}}(\gamma)^r - \chi_{\text{cyclo}}(\gamma))e_1}{(1 - e_1) + (\chi_{\text{cyclo}}(\gamma)^r - 1)e_1} \cdot \xi.$$

This verifies the identity of Proposition 7.

6. Results in the tamely ramified case

We resume our notation and assumptions from Section 4.3. Our first aim in this section is to prove Proposition 44 below which is a yet more explicit reformulation of the identity (16) in Proposition 7. We then prove this identity for $e < p$ and $r = 1$ as well as for $e < p/4$ and $r = 2$. In the isotypic components where $\eta|_{\text{Gal}(K/F(\zeta_p))} = 1$ this can easily be done (for any $r$) using computations similar to those in Section 5.3 with

$$\beta_1 := \text{pr}_{1,1-r}(\beta_{Iw})$$

and $\beta_{Iw}$ defined in (30). The notation here is relative to the base field $K = F$. In any case, the equivariant local Tamagawa number conjecture is known for any $r$ in those isotypic components by [Benois and Berger 2008]. We shall therefore entirely focus on isotypic components with

$$\eta|_{\text{Gal}(K/F(\zeta_p))} \neq 1.$$

In this case we need to verify Equation (16). The main problem is that we do not have any closed formula for a $\Lambda_K$-basis of (the torsion free part of) $A^K_{\psi=1}$. We shall analyze a general basis using Nakayama’s lemma, and to do this we first need to analyze which restrictions are put on a power series

$$a = \sum_n a_n \pi^K_n \in A_K$$

by the condition $\psi(a) = a$.

6.1. Analyzing the condition $\psi = 1$. Proposition 34 below, which is the main result of this subsection, gives the rate of convergence of $a_n \to 0$ as $n \to -\infty$ for $a \in A^K_{\psi=1}$. 

Definition 26. For $n \in \mathbb{N}_0$ and $m \in \mathbb{Z}_{(p)}$ define
\[
b_{m,n} := p^{-1} \sum_{\zeta \in \mu_p} \zeta^m (1 - \zeta^{-1})^n
\]
\[= p^{-1} \text{Tr}_{\mathbb{Q}(\zeta_p) / \mathbb{Q}} \zeta_p^m (1 - \zeta_p^{-1})^n, \quad \text{if } n \geq 1.
\]
Clearly $b_{m,n}$ only depends on $m \pmod{p}$.

Lemma 27. One has $b_{m,n} \in \mathbb{Z}$ and
\[
b_{m,n} = \begin{cases} (-1)^{\overline{m}} \binom{n}{\overline{m}}, & 0 \leq n < p, \\
(-1)^{\overline{m}} \binom{n}{\overline{m}} - (-1)^{\overline{m}} \binom{n}{\overline{m} + p}, & p \leq n < 2p,
\end{cases}
\]
where $0 \leq \overline{m} < p$ is the representative for $m \pmod{p}$. Moreover,
\[
p \left\lceil \frac{n + p - 2}{p - 1} \right\rceil - 1 \mid b_{m,n}
\]
for $n \geq 1$ and hence
\[
p^j \mid b_{m,n}
\]
for $j(p - 1) < n \leq (j + 1)(p - 1)$.

Proof. Formula (32) follows from the binomial expansion of $(1 - \zeta^{-1})^n$ and the fact that
\[
\sum_{\zeta \in \mu_p} \zeta^k = \begin{cases} 0, & p \nmid k, \\
p, & p \mid k.
\end{cases}
\]
In particular $b_{m,0} = 0, 1$ according to whether $p \nmid m$ or $p \mid m$. The different of the extension $\mathbb{Q}(\zeta_p) / \mathbb{Q}$ is $(1 - \zeta_p)^{p-2}$, so we have
\[
\text{Tr}_{\mathbb{Q}(\zeta_p) / \mathbb{Q}} (\zeta_p^m (1 - \zeta_p^{-1})^n) \subseteq p^N \mathbb{Z}
\]
\[
\iff ((1 - \zeta_p)^n) \subseteq \left(p^N (1 - \zeta_p)^{2-p}\right) = ((1 - \zeta_p)^{N(p-1)+2-p})
\]
\[
\iff n \geq N(p - 1) + 2 - p \iff N \leq \frac{n + p - 2}{p - 1}.
\]

Definition 28. Define integers $\beta_{n,j} \in \mathbb{Z}$ by $\beta_{1,j} := \frac{1}{p} \binom{p}{j}$ for $1 \leq j \leq p - 1$ and
\[
\left( \sum_{j=1}^{p-1} \beta_{1,j} x^j \right)^n = \sum_{j=n}^{n(p-1)} \beta_{n,j} x^j.
\]

Proposition 29. An element $a = \sum_i a_i \pi_K^i \in A_K$ lies in $A_K^{\psi=1}$ if and only if for all $N \in \mathbb{Z}$ one has
\[
\sum_{n=0}^{\infty} a_{N+en} \binom{N/e + n}{n} b_{(N/e)+n,n} = \sum_{0 \leq n \leq j \leq n(p-1)} a_{(N+je)/p}^\sigma \binom{N+je}{pe/n} \beta_{n,j} \cdot p^n
\]
with the convention that $a_r = 0$ for $r \notin \mathbb{Z}$. **Equation (33)** holds for all $N \in \mathbb{Z}$ if and only if it holds for all $N \in p\mathbb{Z}$.

**Proof.** This is just comparing coefficients in the identity $p^{-1} \operatorname{Tr}_{B/\varphi(B)}(a) = \varphi(a)$. One has $\varphi(\pi) = (1 + \pi)^p - 1 = \pi^p (1 + p \cdot y)$ with $y = \sum_{j=1}^{p-1} \beta_1 \pi^{-j}$ and hence

$$\varphi(\pi_K) = \pi_K^p \cdot \lambda \cdot (1 + p \cdot y)^{1/e}$$

with $\lambda \in \mu_e$ and $(1 + Z)^{1/e}$ the binomial series. In fact, $\lambda = 1$ since $\varphi(\pi_K) \equiv \pi_K^p \mod p$. Therefore

$$\varphi(\pi_K^m) = \pi_K^{pm} (1 + p \cdot y)^{m/e} = \pi_K^{pm} \sum_{n=0}^{\infty} \left( \frac{m}{e} \right) y^n \cdot p^n$$

$$= \sum_{n=0}^{\infty} \left( \frac{m}{e} \right) \sum_{j=n}^{p-1} \beta_n \pi^{pm- ej} \cdot p^n$$

and the coefficient of $\pi_K^N$ in $\varphi(a) = \sum_m a_m \varphi(\pi_K^m)$ is

$$\sum_{m,n,j,N=pm-ej} a_m \left( \frac{m}{e} \right) \beta_{n,j} \cdot p^n,$$

which is the right-hand side of (33). The conjugates of $\pi$ over $\varphi(B)$ are

$$(1 + \pi) \zeta - 1 = \pi \cdot \zeta \cdot (1 + (1 - \zeta^{-1})\pi^{-1}),$$

hence the conjugates of $\pi_K^m$ are

$$\pi_K^m \cdot \zeta^{m/e} \cdot (1 + (1 - \zeta^{-1})\pi^{-1})^{m/e} = \pi_K^m \cdot \zeta^{m/e} \cdot \sum_{n=0}^{\infty} \left( \frac{m}{e} \right) (1 - \zeta^{-1})^n \pi^{-n}$$

and

$$p^{-1} \operatorname{Tr}_{B/\varphi(B)}(\pi_K^m) = \pi_K^m \sum_{n=0}^{\infty} \left( \frac{m}{e} \right) b_{m/e,n} \pi^{-n} = \sum_{n=0}^{\infty} \left( \frac{m}{e} \right) b_{m/e,n} \pi_K^{m- en},$$

and the coefficient of $\pi_K^N$ in $p^{-1} \operatorname{Tr}_{B/\varphi(B)}(a)$ is the left-hand side of (33). Note here that $B(\zeta)/\varphi(B)$ is totally ramified, so all the conjugates must be congruent modulo $1 - \zeta$.

Denote by $(33)_m$ the equation (33) modulo $p^m$. By Lemma 30 below, $(33)_1$ for all $N \in \mathbb{Z}$ is equivalent to $(33)_1$ for all $N \in p\mathbb{Z}$. We shall show by induction on $m$ that this equivalence holds for all $m$. Suppose $a \in A_K$ satisfies $(33)_{m+1}$ for all $N \in p\mathbb{Z}$. Let $b \in A_K^{\psi=1}$ be a lift of $\tilde{a} \in E_K^{\psi=1}$, which exists by Lemma 32 below, and write $a - b = c \cdot p$. Then $a - b$ satisfies $(33)_{m+1}$ for all $N \in p\mathbb{Z}$, hence $c$ satisfies $(33)_m$ for all $N \in p\mathbb{Z}$. By the induction assumption $c$ satisfies $(33)_m$ for all $N \in \mathbb{Z}$. But then $p \cdot c$ satisfies $(33)_{m+1}$ for all $N \in \mathbb{Z}$, hence so does $a = b + c \cdot p$. \qed
Lemma 30. An element \( a = \sum_i a_i \pi^i_k \in E_K \) lies in \( E^{\psi=1}_K \) if and only if for all \( k \in \mathbb{Z} \) one has
\[
\sum_{n=0}^{p-1} a_{kp+ne}(-1)^n = a^\sigma_k. \tag{34}
\]

Proof. The only nonzero term on the right-hand side of (33)_1 is \( a^\sigma_{N/p} \), corresponding to \( n = j = 0 \), and the nonzero terms on the left-hand side are for \( n \leq p-1 \) by Lemma 27. For \( m \in \mathbb{Z}_{(p)} \) one has
\[
\binom{m}{n} \left( \frac{n}{m} \right) = \frac{m(m-1) \cdots (m-n+1)}{n!} \cdot \frac{n!}{m!(n-m)!} \equiv \begin{cases} 0, & m < n, \\ 1, & m = n, \end{cases}
\]
since for \( m < n \) one of the factors in \( m(m-1) \cdots (m-n+1) \) is divisible by \( p \), whereas for \( m = n \) this product is congruent to \( m! \) modulo \( p \). For \( m > n \) one has \( \binom{m}{n} = 0 \), so \( \binom{m}{n} \binom{n}{m} = 0 \) whenever \( m \neq n \). Using (32) the left-hand side of (33)_1 is
\[
\sum_{n=0}^{p-1} a_{N+en} \binom{m}{n} \left( \frac{n}{m} \right)(-1)^n
\]
for \( m = (N/e) + n \). So the left-hand side vanishes for \( N \notin p\mathbb{Z} \) and is equal to the left-hand side of (34) for \( N = pk \). \( \square \)

For later reference we also record here a more explicit version of (33)_2.

Lemma 31. Let \( H_0 = 0 \) and \( H_n = \sum_{i=1}^n 1/i \) be the harmonic number. Then (33)_2 holds if and only if for all \( k \in \mathbb{Z} \) one has
\[
\sum_{n=0}^{p-1} a_{kp+ne}(-1)^n \left( 1 + \frac{kp}{e} H_n \right) + \sum_{n=p+1}^{2(p-1)} a_{kp+ne}(-1)^{n-p} \cdot p \cdot H_{n-p} \left( 1 + \frac{k}{e} \right) \equiv a^\sigma_k. \tag{35}
\]

Proof. The only nonzero term on the right-hand side of (33)_2 for \( N = kp \) is \( a^\sigma_k \), corresponding to \( n = j = 0 \), since for \( n = 1 \) there is no \( 1 \leq j \leq (p-1) \) with \( p \mid (N+je) = kp+je \). The nonzero terms on the left-hand side are for \( n \leq 2(p-1) \) by Lemma 27. Note that for \( 1 \leq j \leq n < 2p \) only \( j = p \) is divisible by \( p \). So computing modulo \( p^2 \) we have
\[
\binom{kp/e + n}{n} = \prod_{j=1}^n \binom{kp/e + j}{n} = \frac{n! + kp/e \sum_{j=1}^n \frac{n!}{j!}}{n!} + \left( \frac{kp/e}{n!} \right) \sum_{1 \leq j_1 < j_2 \leq n} \frac{1}{j_1 j_2}
\]
\[
\equiv 1 + \frac{kp}{e} H_n + \left( \frac{kp}{e} \right)^2 \sum_{1 \leq j_1 < j_2 \leq n} \frac{1}{j_1 j_2}
\]
\[
\equiv \begin{cases} 1 + \frac{kp}{e} H_n, & n < p, \\ 1 + \frac{k}{e} + \frac{kp}{e} H_{n-p} + \left( \frac{k}{e} \right)^2 \cdot p \cdot H_{n-p}, & p \leq n < 2p. \end{cases}
\]
Here we have used $H_{p-1} \equiv 0 \mod p$ and $\sum_{j=p+1}^{n} 1/j \equiv H_{n-p} \mod p$. By (32) we have

$$b_{(kp/e)+n,n} = \begin{cases} 
(\binom{n}{p})(-1)^n = (-1)^n, & n < p, \\
0, & n = p, \\
(-1)^{n-p}\left(\binom{n}{n-p} - \binom{n}{n}\right), & p < n < 2p 
\end{cases}$$

and

$$\binom{n}{n-p} - \binom{n}{n} = \frac{(p+n-p)(p+n-p-1)\cdots(p+1)}{(n-p)!} - 1 \equiv p \cdot \sum_{j=1}^{n-p} \frac{1}{j}.$$ 

So the summand for $n = p$ vanishes and for $p < n < 2p$ we have

$$\binom{k\cdot p + n}{n} b_{(kp/e)+n,n} \equiv \left(1 + \frac{k}{e} + \frac{kp}{e} H_{n-p} + \left(\frac{k}{e}\right)^2 \cdot p \cdot H_{n-p}\right)(-1)^{n-p} \cdot p \cdot H_{n-p} \equiv (-1)^{n-p}\left(1 + \frac{k}{e}\right) \cdot p \cdot H_{n-p}.$$

**Lemma 32.** The map $A_K^{\psi=1} \to E_K^{\psi=1}$ is surjective.

**Proof.** This follows from the snake lemma applied to

$$\begin{array}{ccccccc}
0 & \to & A_K & \xrightarrow{p} & A_K & \xrightarrow{\psi-1} & E_K & \to & 0 \\
& & \downarrow & & \downarrow & & \downarrow & & \\
0 & \to & A_K & \xrightarrow{p} & A_K & \xrightarrow{\psi-1} & E_K & \to & 0
\end{array}$$

and the fact that $A_K/(\psi - 1)A_K \cong H_{Iw}^2(K, \mathbb{Z}_p(1)) \cong \mathbb{Z}_p$ (see [Cherbonnier and Colmez 1999, remarque II.3.2.1]) is $p$-torsion free. 

**Definition 33.** For $a = \sum_i a_i \pi_K^i \in A_K$ and $\nu \geq 1$ we set

$$l_\nu(a) := \min\{i \mid p^\nu \mid a_i\}.$$ 

In particular

$$l(a) := l_1(a) = v_{\pi_K}(\bar{a})$$

is the valuation of $\bar{a} \in E_K$.

Note that $l(a)$ is independent of a choice of uniformizer for $A_K$, but for $\nu \geq 2$, $l_\nu(a)$ is not.

**Proposition 34.** Let $a \in A_K^{\psi=1}$.

(a) For all $\nu \geq 1$ we have

$$l_\nu(a) \geq -\frac{\nu(p-1)+1}{p} \cdot e.$$
In particular \( l(a) \geq -e \).

(b) If \( l(a) < -e + e(p - 1) \) then

\[
l_2(a) > l(a) - e(p - 1),
\]

while if \( l(a) \geq -e + e(p - 1) \) then \( l_2(a) \geq -e \).

(c) If \( l(a) < -e + 2e(p - 1) \) and \( l_2(a) \geq l(a) - e(p - 1) \) then

\[
l_3(a) > l(a) - 2e(p - 1),
\]

while if \( l(a) \geq -e + 2e(p - 1) \) and \( l_2(a) \geq l(a) - e(p - 1) \) then \( l_3(a) \geq -e \).

Remark 35. Part (b) is a small improvement of part (a) for \( \nu = 2 \) and \( a \) with

\[
l(a) > -(2 - \frac{1}{p})e + e(p - 1),
\]

while part (c) improves (a) for \( \nu = 3 \) and \( a \) with

\[
l(a) > -(3 - \frac{2}{p})e + 2e(p - 1)
\]

and \( l_2(a) \geq l(a) - e(p - 1) \).

Proof. Suppose \( a = \sum_i a_i \pi_K^i \in A_{K}^{\psi=1} \). Part (a) is equivalent to the statement

\[
i < -\frac{\nu(p - 1) + 1}{p} \cdot e \Rightarrow p^\nu | a_i, \tag{36}
\]

which we denote by \((36)_\nu\) if we want to emphasize dependence on \( \nu \). We shall prove \((36)_\nu\) by induction on \( \nu \), the statement \((36)_0\) being trivial. Now assume \((36)_{\nu'}\) for \( \nu' < \nu \) and assume \( p^{\nu'+1} \nmid a_i \) for some

\[
i < -\frac{(\nu + 1)(p - 1) + 1}{p} \cdot e.
\]

We shall show that there is another \( i' < i \) with \( p^{\nu'+1} \nmid a_i \). Hence there are infinitely many \( i < 0 \) with \( p^{\nu'+1} \nmid a_i \) which contradicts the fact that \( a \in A_K \). This proves \((36)_{\nu+1}\).

In order to find \( i' \) we look at (33) for \( N = pi \)

\[
\sum_{n=0}^{\infty} a_{p^i+e} n \left( \frac{p^i}{e} + n \right) b_{(p^i/e)+n,n} = a_i^\sigma + \sum_{1 \leq n \leq p^\lambda e \leq (p-1)} a^\sigma_{i+\lambda e} \left( \frac{1}{e} + \lambda \right) b_{n,p^\lambda} \cdot p^n \tag{37}
\]

and first notice that

\[
p^{\nu+1-n} | a_{i+\lambda e}
\]
for \( n/p \leq \lambda \leq n(p - 1)/p \). This is because of

\[
i + \lambda e < -\frac{(v + 1)(p - 1) + 1}{p} \cdot e + \frac{n(p - 1)}{p} \cdot e = -\frac{(v + 1 - n)(p - 1) + 1}{p} \cdot e
\]

and the induction assumption. Since \( (\frac{i}{n} + \lambda)\beta_{n,p} \) is a \( p \)-adic integer we conclude that \( p^{v+1} \) divides the sum over \( \lambda, n \) in the right-hand side of (37) and hence does not divide the right-hand side of (37).

Considering the left-hand side of (37) we first recall that Lemma 27 implies that

\[
p^j \mid b_{(pi/e)+n,n}
\] (38)

for \( j(p - 1) < n \leq (j + 1)(p - 1) \). For \( n \) in this range we have

\[
pi + ne \leq pi + (j + 1)(p - 1)e < -(vi + 1)(p - 1) + 1e + (j + 1)(p - 1)e \\
= -(v + 1 - j)(p - 1) + 1e + (p - 1)e \\
\leq -\frac{(v + 1 - j)(p - 1) + 1}{p} \cdot e
\] (39)

provided this last inequality holds which is equivalent to

\[
p((v + 1 - j)(p - 1) + 1) - p(p - 1) \geq (v + 1 - j)(p - 1) + 1 \\
\iff (p - 1)((v + 1 - j)(p - 1) + 1) \geq p(p - 1) \\
\iff ((v + 1 - j)(p - 1) + 1) \geq p \\
\iff (v + 1 - j) \geq 1 \iff v \geq j.
\]

So for \( 1 \leq j \leq v \) inequality (39) holds, and the induction assumption implies

\[
p^{v+1-j} \mid a_{pi+ne}.
\]

Using (38) we conclude that \( p^{v+1} \) divides all summands in the left-hand side of (37) except perhaps those with \( n < p \) (corresponding to \( j = 0 \)). Since \( p^{v+1} \) does not divide the right-hand side, it does not divide the left-hand side of (37). So there must be one summand with \( n < p \) not divisible by \( p^{v+1} \) and hence some \( i' := pi + en \) with \( n \leq p - 1 \) such that \( p^{v+1} \nmid a_{i'} \). It remains to remark that

\[
i' = pi + en \leq pi + e(p - 1) < pi - i(p - 1) = i
\] (40)

since \( i < -e \).

To prove (b) we use the same argument. Assuming the existence of

\[
i \leq \min\{l(a) - e(p - 1), -e - 1\}
\]

with \( p^2 \nmid ai \) we find another \( i' < i \) with \( p^2 \nmid ai' \). On the right-hand side of (37), apart from \( ai' \), all summands are divisible by \( p^2 \) (note there are none with \( n = 1 \) since \( \lambda \),
has to be an integer). On the left-hand side, summands for \( n > 2(p-1) \) are divisible by \( p^2 \) by Lemma 27. For \( p \leq n \leq 2(p-1) \) we have, assuming \( l(a) < -e + e(p-1) \),

\[
pi + en \leq (l(a) - e(p-1)) + 2(p-1)e = l(a) + (p-1)l(a) - (p-2)(p-1)e
\]

and therefore \( p \mid a_{pi+en} \). If \( l(a) \geq -e + e(p-1) \) we have

\[
pi + en < p(-e) + 2(p-1)e = -e + e(p-1) \leq l(a)
\]

and again conclude \( p \mid a_{pi+en} \). So all summands on the left-hand side with \( n \geq p \) are divisible by \( p^2 \). Hence some \( i' := pi + en \) with \( n \leq p - 1 \) satisfies \( p^2 \mid a_{i'} \). Moreover, (40) holds since \( i < -e \).

For (c) we use this argument yet another time. Assume

\[
i \leq \min\{l(a) - 2e(p-1), -e - 1\}
\]

and \( p^3 \mid a_i \). On the right-hand side of (37) we need \( p \mid a_{i + \lambda e} \) for \( 2/p \leq \lambda \leq 2(p-1)/p \), i.e., \( \lambda = 1 \). But

\[
i + e \leq \min\{l(a) - 2e(p-1) + e, -1\} < l(a),
\]

so \( p \mid a_{i+e} \). Assume first \( l(a) < -e + 2e(p-1) \). On the left-hand side we have for \( p \leq n \leq 2(p-1) \)

\[
pi + en \leq p(l(a) - 2e(p-1)) + 2(p-1)e
\]

\[
= l(a) - e(p-1) + (p-1)l(a) + e(p-1) - (2p-2)(p-1)e
\]

\[
< l(a) - e(p-1) + (p-1)(-e + 2e(p-1)) - (2p-3)(p-1)e
\]

\[
= l(a) - e(p-1) \leq l_2(a)
\]

and therefore \( p^2 \mid a_{pi+en} \). For \( 2p - 1 \leq n \leq 3(p-1) \) we just add \( (p-1)e \) to this last estimate to conclude

\[
pi + en \leq p(l(a) - 2e(p-1)) + 3(p-1)e
\]

\[
< l(a) - e(p-1) + e(p-1) = l(a)
\]

and hence \( p \mid a_{pi+en} \). Now assume \( l(a) \geq -e + 2e(p-1) \). For \( p \leq n \leq 2(p-1) \) we have

\[
pi + en \leq p(-e) + 2(p-1)e \leq l(a) - e(p-1) \leq l_2(a)
\]

and therefore \( p^2 \mid a_{pi+en} \). For \( 2p - 1 \leq n \leq 3(p-1) \) we again add \( (p-1)e \) to this last estimate to conclude \( pi + en < l(a) \) and \( p \mid a_{pi+en} \). As before we conclude that, for some \( i' := pi + en \) with \( n \leq p - 1 \), we have \( p^3 \mid a_{i'} \). Moreover (40) holds since \( i < -e \).
Before drawing consequences of Proposition 34 we make the following definition.

**Definition 36.** Let \( \varpi \) be the uniformizer of \( K \) given by
\[
\varpi = \sqrt[\varphi p - 1 = \varphi^{-1}(\pi_K)|_{t=0}}
\]
and denote by \( v_{\varpi} \) the unnormalized valuation of the field \( K \), i.e.,
\[
v_{\varpi}(p) = e(p - 1).
\]
For \( a \in B_K^{+,1} \) define
\[
v_{\varpi}(a) := v_{\varpi}(\varphi^{-1}(a)|_{t=0}).
\]

**Corollary 37.** For all \( a \in A_K^{\psi=1} \) the series \( \varphi^{-1}(a) \) converges, i.e., \( A_K^{\psi=1} \subseteq B_K^{+,1} \).

**Proof.** By (a) we have \( p^\nu | a_i \) for
\[
-(\nu + 1)(p - 1) + 1 < i < -\nu(p - 1) + 1.
\]
and hence
\[
v_p(a_i) \geq v \geq -\frac{ip + e}{e(p - 1) - 1}
\]
and
\[
v_{\varpi}(a_i \varpi^i) \geq -(ip + e) - e(p - 1) + i = -(p - 1)i - pe.
\]
This implies
\[
\lim_{i \to -\infty} v_{\varpi}(a_i \varpi^i) = \infty
\]
and hence the series \( \sum_{i \in \mathbb{Z}} a_i \varpi^i \) converges in \( K \subseteq \hat{\mathbb{Q}}_p \). By [Colmez 1999, proposition II.25] this implies that \( \varphi^{-1}(a) \) converges in \( B_{dR} \). \(\square\)

**Proposition 38.** For each \( a \in E_K^{\psi=1} \) we have \( l(a) \geq -e \). If \( l(a) > -e \) then \( l(a) \not\equiv -e \mod p \). Conversely, for each \( c \in k^\times \) and \( n \in \mathbb{Z} \) with
\[
-e < n \not\equiv -e \mod p
\]
there is an element \( a \in E_K^{\psi=1} \) with \( l(a) = n \) and leading coefficient \( c \).

**Proof.** That \( l(a) \geq -e \) is Proposition 34(a). Assume that \( l(a) > -e \) and \( l(a) \equiv -e \mod p \). Then \( l(a) = kp + (p - 1)e \) for some \( k \in \mathbb{Z} \) and
\[
k = \frac{l(a) - (p - 1)e}{p} = l(a) - \left(1 - \frac{1}{p}\right)(l(a) + e) < l(a),
\]
so we have \( a_k = 0 \). Further, \( a_{kp+ie} = 0 \) for \( i = 0, \ldots, p - 2 \) since \( kp + ie < l(a) \). Hence there is only one nonzero term in (34) which gives a contradiction.
To show the second part one can solve (34) by an easy recursion. Alternatively, Proposition 10 implies that $\nabla \log(a) \in E_K^{\psi=1}$ for any $a \in E_K^\times$. Now compute

$$\nabla \log(1 + c\pi_K^n) = \frac{\nabla(1 + c\pi_K^n)}{1 + c\pi_K^n} = \frac{cn/e \cdot (\pi_K^n - e + \pi_K^n)}{1 + c\pi_K^n} = \frac{cn}{e} \cdot \pi_K^n - e + \ldots$$

and note that for $p \nmid n$ one can produce any leading coefficient. $\square$

**Remark 39.** Elements $a \in E_K^{\psi=1}$ with $l(a) = -e$ exist, e.g.,

$$\nabla \log(\pi^j) = j \cdot \pi^{-1} + j \cdot \pi_K^{-e} + j,$$

but their leading coefficient is restricted to elements in $\mathbb{F}_p$.

**Corollary 40.** If $a \in A_K^{\psi=1}$ and

$$l(a) < -e + e(p - 1),$$

we have $v_{\sigma}(a) = l(a)$.

**Proof.** Since $v_{\sigma}(a_{l(a)\sigma^{l(a)}}) = l(a)$ we need to show

$$v_{\sigma}(a_i \sigma^i) > l(a)$$

for $i \neq l(a)$. This is clear for $i > l(a)$, and also for

$$l(a) - e(p - 1) < i < l(a)$$

since in that range $p \mid a_i$ and so $v_{\sigma}(a_i \sigma^i) \geq e(p - 1) + i > l(a)$. For

$$l(a) - 2e(p - 1) < i \leq l(a) - e(p - 1)$$

we have $p^2 \mid a_i$ by part (b) and hence $v_{\sigma}(a_i \sigma^i) \geq 2e(p - 1) + i > l(a)$. Finally for

$$i \leq l(a) - 2e(p - 1) < -e - e(p - 1) = -ep < -2e$$

we have by (41)

$$v_{\sigma}(a_i \sigma^i) \geq -(p - 1)i - pe > (p - 1)2e - pe = (p - 2)e > l(a),$$

using the assumption on $l(a)$. $\square$

In order to study $v_{\sigma}(a)$ for $a \in A_K^{\psi=1}$ with $l(a) > -e + e(p - 1)$ we need to use Lemma 31. The next proposition will show that $v_{\sigma}(a)$ cannot only depend on $l(a)$ in this case. In the situation of Proposition 41(b) one can have $v_{\sigma}(a) = l(a)$ but for any $b \in A_K^{\psi=1}$ with $l(b) < l(a) - e(p - 1)$ and $p^2 \nmid a_{l(b)} + pb_{l(b)}$ one has

$$l(a + pb) = l(a), \quad v_{\sigma}(a + pb) \leq l(b) + e(p - 1) < l(a) = v_{\sigma}(a).$$
Proposition 41. Let \( a' \in A^\psi_K \) with
\[
l(a') = \mu p - e + e(p - 1)
\]
for some \( \mu \in \mathbb{Z} \) with \( 1 \leq \mu < \frac{e(p-1)}{p} \).

(a) There exists \( a \equiv a' \mod p \) with
\[
l_2(a) \geq \mu p - e = l(a) - e(p - 1).
\]

(b) For \( a \) as in (a) we have \( v_\sigma(a) \geq l(a) \) with equality if \( p \nmid \mu - e \). This last condition is automatic for \( e < p \).

Proof. First note that \( l_2(a') \geq -e \) by Proposition 34(b). If \( l_2(a') = -e \) then Equation (35) for \( k := -e \) reads
\[
a'_{-e}^{-} \equiv a'_{kp + e(p-1)} = a'_{-e}
\]

since \( i = kp + en < l_2(a') \) for \( n < p - 1 \) and \( i = kp + en \leq -e + e(p - 1) < l(a') \) for \( p + 1 \leq n \leq 2(p - 1) \). Hence \( a'_{-e}^{-}/p \mod p \in \mathbb{F}_p \). Adding an element \( pb \) to \( a' \), where \( b \) with \( l(b) = -e \) is as in Remark 39, we can assume that \( l_2(a') > -e \). More generally, as long as \( l_2(a') < l(a') \), we can add elements \( pb \) to \( a' \) whose existence is guaranteed by Proposition 38 and increase \( l_2(a') \) until \( l_2(a') \) is not one of the possible \( l(b) \), i.e.,
\[
l_2(a') = \mu' p - e = (\mu' - e) p + (p - 1)e
\]

for some \( \mu' \geq 1 \). Equation (35) for \( k := \mu' - e \) then reads
\[
0 \equiv a'_{kp + e(p-1)} + \sum_{n=0}^{2(p-1)} a'_{kp + ne} \cdot (-1)^{n-p} \cdot p \cdot H_{n-p} \left( 1 + \frac{k}{e} \right)
\]

since \( i = kp + en < l_2(a') \) for \( n < p - 1 \) and also \( i = k < l_2(a') \), so \( a'_i \equiv 0 \) for those \( i \). If \( \mu' < \mu \) we have for \( p + 1 \leq n \leq 2(p - 1) \)
\[
kp + ne < (\mu - e) p + 2(p - 1)e = l(a')
\]

and hence \( p \mid a'_{kp + ne} \). So if \( \mu' < \mu \) then \( a'_{kp + e(p-1)} \) is the only nonzero term in (42) and we arrive at a contradiction. Therefore \( \mu' \geq \mu \) and we have found our \( a' \), or otherwise we arrive at an \( a' \) with \( l_2(a') = l(a') \). In either case this proves part (a).

Equation (42) for \( k := \mu - e \) gives
\[
0 \equiv a_{kp + e(p-1)} + a_{l(a)} \cdot (-1) \cdot p \cdot H_{p-2} \left( 1 + \frac{\mu - e}{e} \right)
\]

\[
= a_{kp + e(p-1)} - a_{l(a)} \cdot p \cdot \frac{\mu}{e} \mod p^2
\]

(43)
since \( p \mid a_{kp + ne} \) for \( kp + ne < kp + 2(p - 1)e = l(a) \). Note also

\[
H_{p-2} = H_{p-1} - \frac{1}{p-1} \equiv 0 - (-1) = 1 \pmod{p}.
\]

For part (b) we need to show that \( v_{\sigma}(a_i \sigma^i) \geq l(a) \) for all \( i \in \mathbb{Z} \) (and compute the sum over those \( i \) for which there is equality). As in the proof of Corollary 40 for \( i > l(a) \) and \( l(a) - e(p - 1) < i < l(a) \) we obviously have \( v_{\sigma}(a_i \sigma^i) > l(a) \). By (43) we have

\[
a_{l(a)-e(p-1)} \sigma^{l(a)-e(p-1)} + a_{l(a)} \sigma^{l(a)} \equiv \left( \frac{p \mu}{\chi \sigma e(p-1) e} + 1 \right) a_{l(a)} \sigma^{l(a)} \\
= \left( -\frac{\mu}{e} + 1 \right) a_{l(a)} \sigma^{l(a)} + O(\sigma^{l(a)+1}) \tag{44}
\]

since

\[
\sigma^{e(p-1)} = (\zeta_p - 1)^{p-1} \equiv -p \pmod{(\zeta_p - 1)^p}.
\]

So if \( p \nmid -(\mu/e) + 1 \) this is the leading term of valuation \( l(a) \). For

\[
l(a) - 2e(p - 1) < i < l(a) - e(p - 1),
\]

since \( l_2(a) \geq l(a) - e(p - 1) \) by part (a), we have \( p^2 \mid a_i \) and hence \( v_{\sigma}(a_i \sigma^i) \geq 2e(p - 1) + i > l(a) \). For

\[
l(a) - 3e(p - 1) < i \leq l(a) - 2e(p - 1)
\]

we have \( p^3 \mid a_i \) by (c) of Proposition 34 and hence \( v_{\sigma}(a_i \sigma^i) \geq 3e(p - 1) + i > l(a) \). Finally for

\[
i \leq l(a) - 3e(p - 1) < -e - e(p - 1) = -ep
\]

we have by (41)

\[
v_{\sigma}(a_i \sigma^i) \geq -(p - 1)i - pe > (p - 1)pe - pe = (p - 2)pe \geq (2p - 3)e > l(a)
\]

using the assumption on \( l(a) \). \( \square \)

### 6.2. Isotypic components.

We introduce some notation for isotypic components. Recall that

\[
G \cong \Sigma \ltimes \Delta
\]

with \( \Sigma \) cyclic of order \( f \) and \( \Delta \) cyclic of order \( e(p - 1) \). For any \( \Sigma \)-orbit \([\eta]\) we define the idempotent

\[
e_{[\eta]} = \sum_{\eta' \in \overline{\Sigma}_{\eta}} e_{\chi} \in \mathbb{Z}_p[G],
\]
where the irreducible characters \( \chi = ([\eta], \eta') \) of \( G \) are parametrized as in Section 3. For any \( \mathbb{Z}_p[G] \)-module \( M \) its \([\eta]\)-isotypic component
\[
M^{[\eta]} := e_{[\eta]}M
\]
is again a \( \mathbb{Z}_p[G] \)-module. The \( \Sigma \)-orbit
\[
[\eta] = \{ \eta, \eta^p, \eta^{p^2}, \ldots, \eta^{p^{n_f-1}} \} = \{ \eta_0^{n_1}, \ldots, \eta_0^{n_f} \}
\]
corresponds to an orbit \( \{ n_1, \ldots, n_{f_\eta} \} \subseteq \mathbb{Z}/e(p-1)\mathbb{Z} \) of residue classes modulo \( e(p-1) \) under the multiplication-by-\( p \) map, i.e., we have \( n_{i+1} \equiv n_i p \mod e(p-1) \) where we view the index \( i \) as a class in \( \mathbb{Z}/f_\eta\mathbb{Z} \). We shall use the notation
\[
[\eta] = \{ n_1, \ldots, n_{f_\eta} \}
\]
to denote both the orbit of residue classes in \( \mathbb{Z}/e(p-1)\mathbb{Z} \) and the orbit of characters. By (21) the group
\[
\Delta_e := \text{Gal}(K/F(\zeta_p))
\]
acts on \( \sqrt[p]{\zeta_p - 1} = \varphi^{-1}(\pi_K)|_{t=0} \) via the character \( \eta_0 \) defined in Section 3 and acts on \( \pi_K \) via \( \eta_0^p \). The \( [\eta] = \{ n_1, \ldots, n_{f_\eta} \} \)-isotypic component of the \( \mathbb{Z}_p[\Sigma \ltimes \Delta_e] \)-module \( A_K \) is
\[
\left\{ a = \sum a_n \pi_K^n \bigg| a_n = 0 \quad \text{for} \quad n \mod e \notin \{ n_1, \ldots, n_{f_\eta} \} \right\},
\]
but \( A^{[\eta]}_K \) is much harder to describe since \( \pi_K \) is not an eigenvector for the full group \( \Delta \). However, there is the following fact about leading terms.

**Lemma 42.** Fix \( \nu \geq 1, a = \sum_j a_j \pi_K^j \in A_K \) and denote by \( e_\eta \in \mathcal{O}_F[\Delta] \) the idempotent for \( \eta = \eta_0^n \). If
\[
p \cdot l_\nu(a) \equiv n \mod e(p-1),
\]
then
\[
l_\nu(e_\eta^\nu a) = l_\nu(a)
\]
and the leading coefficients modulo \( p^\nu \) of \( e_\eta^\nu a \) and \( a \) agree. If \( a = e_\eta^\nu a \) is an eigenvector for \( \Delta \) then (46) holds.

**Proof.** Denote by
\[
\omega : \Delta \to \text{Gal}(F(\zeta_p)/F) \to \mathbb{Z}_p^\times
\]
the Teichmüller character. For \( \delta \in \Delta \) we have
\[
\delta(\pi_K) = ((1 + \pi)^{\omega(\delta)} - 1)^{1/e} = \left( \sum_{i=1}^{\infty} \frac{\omega(\delta)}{i} \pi^i \right)^{1/e}
\]
\[
= \lambda(\delta) \pi_K \left( 1 + \sum_{i=2}^{\infty} \frac{1}{\omega(\delta) i} \left( \frac{\omega(\delta)}{i} \right) \pi^{i-1} \right)^{1/e}
\]
where \( \lambda(\delta) \in \mu_e(p-1) \) satisfies \( \lambda(\delta)^e = \omega(\delta) \) and \( (1 + Z)^{1/e} \) denotes the usual binomial series. Applying \( \varphi^{-1}|_{\ell=0} \) we find
\[
\delta(\sqrt[1/p]{\zeta_p - 1}) \equiv \lambda(\delta)^{1/p} \cdot \sqrt[1/p]{\zeta_p - 1} \mod \varphi^2
\]
and since \( \sqrt[1/p]{\zeta_p - 1} \equiv \epsilon(p^{-1}/\sqrt[1/p]{-p}) \mod \varphi^2 \) we obtain \( \lambda(\delta) = \eta_0(\delta)^p \). In particular, for any \( a \in A_K \)
\[
\delta(a) \equiv \eta_0(\delta)^{p \cdot l_v(a)} \cdot a l_v(a) \cdot \pi_K^{l_v(a)} \mod (p^\nu, \pi_K^{l_v(a)+1})
\]
and
\[
e_q a = \frac{1}{e(p-1)} \sum_{\delta \in \Delta} \eta^{-1}(\delta) \delta(a) \equiv \frac{1}{e(p-1)} \sum_{\delta \in \Delta} \eta_0(\delta)^{p \cdot l_v(a) - n} \cdot a l_v(a) \cdot \pi_K^{l_v(a)}
\]
\[
\equiv \begin{cases} 
\delta l_v(a) \cdot \pi_K^{l_v(a)} & \text{if } p \cdot l_v(a) \equiv n \mod e(p-1), \\
0 & \text{if } p \cdot l_v(a) \not\equiv n \mod e(p-1),
\end{cases}
\]
where the congruences are modulo \( (p^\nu, \pi_K^{l_v(a)+1}) \). This implies both statements in the lemma.

\( \square \)

**Remark 43.** With the notation introduced in this section we have
\[
e[\eta] = \sum_{i=1}^{f_{\eta}} e_{\eta^i}.
\]

6.3. **The main result.** We view \( \Sigma \) as a subgroup of \( G \) such that \( \epsilon(p^{-1}/\sqrt[1/p]{-p}) \in K^\Sigma \), where \( \epsilon(p^{-1}/\sqrt[1/p]{-p}) \) is the choice of root corresponding to our choice of root \( \pi_K \) of \( \pi \). Then the \( \mathbb{Z}_p[\Sigma] \)-algebra \( \mathbb{Z}_p[G] \) is finite free of rank \( e(p-1) \). For each choice of \( \eta \) the \([\eta]\)-isotypic component of \( \mathbb{Z}_p[G] \) is free of rank \( f_{\eta} \) over \( \mathbb{Z}_p[\Sigma] \) and for each \( \eta \neq \omega \) the \([\eta]\)-isotypic component
\[
(A_K^{\psi=1}(1))_{[\eta]}
\]
of \( A_K^{\psi=1}(1) \) is free of rank \( f_{\eta} \) over \( \mathbb{Z}_p[\Sigma][[\gamma_1 - 1]] \). Write
\[
[\eta] = \{n_1, \ldots, n_{f_{\eta}}\} = [n_1] \subseteq \mathbb{Z}/e(p-1)\mathbb{Z}
\]
and pick representatives \( n_i \in \mathbb{Z} \) with
\[
0 < n_i < e(p-1), \quad i = 1, \ldots, f_{\eta}.
\]
Note that our running assumption \( \eta|_{\Delta_e} \neq 1 \) implies \( e \nmid n_i \).

**Proposition 44.** Fix \( \eta|_{\Delta_e} \neq 1 \) and let \( \{a_i | i = 1, \ldots, f_{\eta}\} \) be a \( \mathbb{Z}_p[\Sigma][[\gamma_1 - 1]] \)-basis of \( (A_K^{\psi=1}(1))_{[\eta]} \). Let \( n_{i,r} \) be representatives for the residue classes
\[
[n_1 - re] \subseteq \mathbb{Z}/e(p-1)\mathbb{Z}
\]
with
\[ 0 < n_{i,r} < e(p - 1) \]
indexed such that \( n_i - re \equiv n_{i,r} \mod e(p - 1) \). Consider the two \( \mathbb{Z}_p[\Sigma] \)-lattices
\[
L_r := \bigoplus_{i=1}^{f_n} \mathbb{Z}_p[\Sigma] \cdot (\nabla^r - 1)(\sqrt[p]{\xi_p} - 1)
\]
and
\[
\mathcal{O}_K^{n_1-re} = \bigoplus_{i=1}^{f_n} \mathcal{O}_F \cdot (\sqrt[p]{1})^{n_{i,r}}
\]
in the \([n_1-re]\)-isotypic component
\[
K^{n_1-re} = \bigoplus_{i=1}^{f_n} F \cdot (\sqrt[p]{1})^{n_{i,r}} = \bigoplus_{i=1}^{f_n} F \cdot (\sqrt[p]{1})^{n_{i-re}}
\]
of \( K \). Then the conjunction of (16) in Proposition 7 for \( \chi = ([n_1-re], \eta') \) over all \( \eta' \) holds if and only if \( L_r \) and \( \mathcal{O}_K^{n_1-re} \) have the same \( \mathbb{Z}_p[\Sigma] \)-volume, i.e.,
\[
\text{Det}_{\mathbb{Z}_p[\Sigma]} L_r = \text{Det}_{\mathbb{Z}_p[\Sigma]} \mathcal{O}_K^{n_1-re}
\]
inside \( \text{Det}_{\mathbb{Z}_p[\Sigma]} K^{n_1-re} \).

Proof. Let \( \alpha \) be a \( \Lambda_K e_{[n_1]} \)-basis of \( (A_{\mathbb{Z}_p}[1])^{[n_1]} \). Then
\[
\beta_{1w} := (\text{Exp}^*_{\mathbb{Z}_p})^{-1}(\alpha)
\]
is a \( \Lambda_K e_{[n_1]} \)-basis of \( H^1_{1w}(K, \mathbb{Z}_p(1))^{[n_1]} \) and the element
\[
\beta = \text{pr}_{1,1-r}(\beta_{1w})
\]
of Corollary 12 is a \( \mathbb{Z}_p[G]e_{[n_1-re]} \)-basis of \( (H^1(K, \mathbb{Z}_p(1-r))/\text{tor})^{[n_1-re]} \). This follows from the fact that the isomorphism \( \text{pr}_{1,1-r} \) of Lemma 8 is not \( \Lambda_K \)-linear but \( \Lambda_K - \kappa_r \)-semilinear, where \( \kappa_j \) is the automorphism of \( \Lambda_K \) given by \( g \mapsto g \chi_{\text{cyclo}}(g)^j \) for \( g \in G \times \Gamma_K \). Theorem 9 and Proposition 13 imply
\[
\exp^*_{\mathbb{Q}_p(r)}(\beta) = \frac{1}{(r-1)!} \left( \frac{d}{dt} \right)^{r-1} p^{-1} \varphi^{-1}(\alpha)|_{t=0}
\]
\[= \frac{p^{-r}}{(r-1)!} \left( \nabla^{r-1} \alpha_{i}^{\sigma-1} \right)(\sqrt[p]{\xi_p} - 1).
\]
Hence the \( \mathbb{Z}_p[G]e_{[n_1-re]} \)-lattice
\[
\mathbb{Z}_p[G] \cdot (r-1)! \cdot p^{r-1} \cdot \exp^*_{\mathbb{Q}_p(r)}(\beta) \subset K^{n_1-re}
\]
is free over \( \mathbb{Z}_p[\Sigma] \) with basis
\[
(r-1)! \cdot p^{r-1} \cdot \frac{p^{-r}}{(r-1)!} \left( \nabla^{r-1} \alpha_{i}^{\sigma-1} \right)(\sqrt[p]{\xi_p} - 1) = p^{-1} \cdot (\nabla^{r-1} \alpha_{i}^{\sigma-1})(\sqrt[p]{\xi_p} - 1),
\]
where $i = 1, \ldots, f_\eta$. Now the conjunction of (16) for $\chi = ([n_1 - re], \eta')$ over all $\eta'$ is equivalent to the statement that the lattice (48) and the $[n_1 - re]$-isotypic component of the inverse different

$$(\sqrt[\psi]{\zeta_p - 1})^{-(e(p-1)-1)} \mathcal{O}_K$$

have the same $\mathbb{Z}_p[\Sigma]$-volume. Since $e \nmid n_1$ we have

$$(\sqrt[\psi]{\zeta_p - 1})^{-(e(p-1)-1)} \mathcal{O}_K^{[n_1-re]} = (p^{-1} \mathcal{O}_K)^{[n_1-re]}$$

and the statement follows. □

### 6.4. Proof for $r = 1, 2$ and small $e$

We retain the notation of the previous section. As in Proposition 24 denote by $\xi$ a $\mathbb{Z}_p[\Sigma]$-basis of $\mathcal{O}_F$.

**Proposition 45.** There exists a $\mathbb{Z}_p[\Sigma][\gamma_1 - 1]$-basis

$$\alpha_i = \xi \cdot \pi^l(\alpha_i) + \cdots \in A^{\psi = 1}_K, \quad i = 1, \ldots, f_\eta$$

of $(A^{\psi = 1}_K)^{[n_1-e]}$ with

$$l(\alpha_i) = \begin{cases} n_i - e & \text{if } p \nmid n_i, \\ n_i - e + e(p-1) & \text{if } p \mid n_i. \end{cases}$$

**Proof.** By Nakayama’s lemma it suffices to find a $\mathbb{F}_p[\Sigma]$-basis for

$$(A^{\psi = 1}_K)^{[n_1-e]}/(p, \gamma_1 - 1) \cong (A^{\psi = 1}_K/(p, \gamma_1 - 1))^{[n_1-e]}.$$

By Lemma 32 we have $A^{\psi = 1}_K/pA^{\psi = 1}_K = E^{\psi = 1}_K$. By Proposition 38 (reductions mod $p$ of) elements $\alpha_i$ as described in Proposition 45 exist in $E^{\psi = 1}_K$. By projection and Lemma 42 we can also assume that they are in the $[n_1-e]$-isotypic component. Let $a'$ be a nonzero $\mathbb{Z}_p[\Sigma]$-linear combination of the $\alpha_i$ and assume

$$a' \equiv (\gamma_1 - 1)a \mod p$$

for some $a \in A^{\psi = 1}_K$. By Lemma 46 below we have $l(a') \geq -e + e(p-1)$. Since $l(a') = l(\alpha_i)$ for some $i$, this implies

$$l(a') \equiv -e + e(p-1) \equiv -2e \mod p.$$
Lemma 46. For $a \in E_{K}^{\psi=1}$ with $l(a) = jp^{\kappa}$ with $p \nmid j$ we have

$$l((\gamma_1 - 1)a) = (j + e(p - 1))p^{\kappa}.$$  

In particular

$$l((\gamma_1 - 1)a) \geq l(a) + e(p - 1)$$

with equality if and only if $p \nmid l(a)$, and

$$l((\gamma_1 - 1)a) \geq -e + e(p - 1)$$

for all $a \in E_{K}^{\psi=1}$.

Proof. Since $\chi_{\text{cyclo}}(\gamma_1) = 1 + p$ we find from (20) that (in $E_{K}$)

$$\gamma_1(\pi) = \pi + \pi p + \pi p^{+1}$$

and hence for $n = jp^{\kappa}$

$$(\gamma_1 - 1)\pi_{K}^{n} = (\pi + \pi p + \pi p^{+1})^{n/e} - \pi^{n/e} = \pi_{K}^{n}\left((1 + \pi p^{1} + \pi p^{p+1})^{n/e} - 1\right)$$

$$= \pi_{K}^{n}((1 + \pi^{p}(p - 1) + \pi^{p+1})^{j/e} - 1)$$

$$= \frac{j}{e} \cdot \pi^{n+ep^{1}(p - 1)} + \ldots$$

and this is indeed the leading term since $p \nmid j$. The last assertion follows from Proposition 34(a). \qed

Proposition 47. If $e < p$, the identity (47) holds for $r = 1$.

Proof. We first remark that for each $i$ we have

$$v_{\omega}(\alpha_i) = l(\alpha_i) = \begin{cases} n_i - e & \text{if } p \nmid n_i, \\ n_i - e + e(p - 1) & \text{if } p \mid n_i \end{cases}$$

by Corollary 40 and Proposition 41. Note that there is at most one $n_i$, $n_1$ say, with

$$0 < n_1 \leq e - 1$$

since all the $n_i$ lie in the same residue class modulo $p - 1$ and $e \leq p - 1$. Then

$$n_2 = pn_1 \leq ep - p < ep - e = e(p - 1)$$

and conversely, $p \mid n_2$ if and only if $0 < n_1 := n_2/p \leq e - 1$. For all other $i$ we have $n_i - e = n_{i,1}$. So if no $n_i - e$ is negative then

$$q_i := \alpha_{i}^{\sigma_{-1}}(\sqrt[p]{\zeta_{p} - 1}) \in K$$

is already a basis of $O_{K}^{[n_1-e]}$. Otherwise

$$p \cdot q_1, p^{-1} \cdot q_2, q_3, \ldots, q_{f_{n}}$$
is a basis of $O_K^{[n_1-e]}$. Since $L_1$ is the span of the $q_i$ the statement follows. □

**Remark 48.** Although not covered by Proposition 2, it is in fact true that the equivariant local Tamagawa number conjecture for $r = 1$ is equivalent to (47) for $r = 1$ and so Proposition 47 proves this conjecture for $e < p$. However, for $r = 1$ one can give a direct proof without any assumption on $e$ other than $p \nmid e$ by studying the exponential map instead of the dual exponential map. Since the exponential power series gives a $G$-equivariant isomorphism

$$\exp : p \cdot O_K \cong 1 + p \cdot O_K,$$

the (equivariant) relative volume of $\exp(O_K)$ and $(O_K^x)^{\wedge} \subseteq H^1(K, \mathbb{Z}_p(1))$ can be easily computed. For more work on the case $r = 1$, see [Bley and Cobbe 2016] and references therein.

To prepare for the proof of Proposition 51 below we need to compute $v_\varpi(\nabla a_i)$, i.e., prove the analogues of Corollary 40 and Proposition 41 for $\nabla a \in A_{K_1}^\psi = p$.

**Lemma 49.** Assume $e < p/2$. For $a \in A_{K_1}^\psi = 1$ with

$$p \nmid l(a) < -e + e(p - 1)$$

or with

$$l(a) = \mu p - e + e(p - 1)$$

and chosen as in Proposition 41(a) we have

$$v_\varpi(\nabla a) = l(\nabla a) = l(a) - e.$$

**Proof.** Since

$$\nabla \pi_K^j = \frac{j}{e} \pi_K^{j-e} + \frac{j}{e} \pi_K^j,$$  \hfill (50)

it is clear that $l(\nabla a) = l(a) - e$ if $p \nmid l(a)$. To compute $v_\varpi(\nabla a)$, note that from the proof of Corollary 40 we already know

$$v_\varpi(a_j \varpi^j) > l(a)$$

for $j \neq l(a)$. But this implies

$$v_\varpi\left(a_j \frac{j}{e} \varpi^{j-e}\right) > l(a) - e, \quad v_\varpi\left(a_j \frac{j}{e} \varpi^j\right) > l(a) > l(a) - e$$  \hfill (51)

for $j \neq l(a)$. This finishes the proof for the case $p \nmid l(a) < -e + e(p - 1)$. If

$$l(a) = \mu p - e + e(p - 1)$$

then recall from the proof of Proposition 41(b) that we had to compute modulo $p^2$ and there were two terms in (44) with valuation $l(a)$ arising from $j = l(a)$ and
Moreover we can choose a with any leading coefficient. Normalizing the leading coefficient to be $\xi$ (as in the $\alpha_i$) we have

$$a \equiv \xi \cdot \frac{\mu p}{e} \cdot \pi_K^{l(a)-e(p-1)} + \cdots + \xi \cdot \pi_K^{l(a)} + \cdots \mod p^2$$

and

$$\nabla a \equiv \xi \cdot \frac{\mu p}{e} \cdot \frac{\mu p - e}{e} \cdot \pi_K^{l(a)-e(p-1)} + \cdots + \xi \cdot \frac{l(a)}{e} \cdot \pi_K^{l(a)-e} + \cdots \mod p^2$$

and hence

$$\frac{\mu p}{e} \cdot \frac{\mu p - e}{e} \cdot \pi_K^{l(a)-e(p-1)} + \frac{l(a)}{e} \cdot \pi_K^{l(a)-e}$$

$$\equiv \left( -\frac{\mu}{e} \cdot \frac{\mu p - e}{e} + \frac{l(a)}{e} \right) \cdot \pi_K^{l(a)-e} \mod p^2.$$}

Computing the leading coefficient modulo $p$ we find

$$\left( \frac{\mu}{e} + \frac{-2e}{e} \right) = \frac{\mu}{e} - 2,$$

which is divisible by $p$ if and only if $p \mid \mu - 2e$. Since $e < p/2$ we have

$$-p < -2e < \mu - 2e < \frac{e(p-1)}{p} - 2e = \left( -1 - \frac{1}{p} \right)e < 0$$

and hence $p \mid \mu - 2e$. In the proof of Proposition 41(b) we showed $v_{\pi} (a_j \pi^j) > l(a)$ for $j \neq l(a), l(a) - e(p-1)$ and as above this implies that the corresponding terms in $\nabla a$ all have valuation larger than $l(a) - e$.

We handle the case $p \mid l(a)$ in a separate lemma. Similar to Proposition 41 we need to compute modulo $p^2$.

**Lemma 50.** Assume $e < p/4$ and $0 < \mu p < -e + e(p-1)$. Then there exists $a \in (A_K^{\psi})^{\mu p}$ with $l(a) = \mu p$ and

$$v_{\pi} (\nabla a) = l(\nabla a) = \mu p - e + e(p-1).$$

Moreover we can choose $a$ with any leading coefficient.

**Proof.** The statement about the leading coefficient will be clear from the proof, so to alleviate notation we take the leading coefficient to be 1. First we can find $a' \in A_K^{\psi}$ with

$$a' \equiv \pi_K^{\mu p} - \pi_K^{\mu p+e(p-1)} + \cdots \mod p^2,$$

i.e., with $a'_i \equiv 0$ for all $i < \mu p + e(p-1)$ and $i \neq \mu p$. To see this, first note that (35) is satisfied for $k = \mu$ since $H_{p-1} \equiv 0 \mod p$ (and we take $a'_{\mu p+ne}$ arbitrary but divisible by $p$ for $n = p+1, \ldots, 2(p-1)$). In any Equation (35) with index $k < \mu$ the coefficient $a'_{\mu p}$ does not occur on the left-hand side since $kp + ne$ is a multiple of $p$ only for $n = 0$ among $n \in \{0, \ldots, p-1, p+1, \ldots, 2(p-1)\}$. On the right-hand
side we always have $a'_k \equiv 0$ since $k < \mu < \mu p$. Similarly, the coefficient $a'_{\mu p + e(p-1)}$ does not occur on the left-hand side for $k < \mu$ since $kp + ne = \mu p + e(p-1)$ implies $n \equiv -1 \pmod{p}$, i.e., $n = p - 1$. So the fact that $a'_i \neq 0$ for $i = \mu p$, $\mu p + e(p-1)$ forces no further nonzero terms in equations with index $k < \mu$. Equations (35) with index $k > \mu$ can always be satisfied inductively by adjusting the variable $a'_{kp+(p-1)e}$ since $a'_{kp+(p-1)e}$ does not occur in any equation with index $k' < k$.

With the notation introduced in Section 6.2 set

$$a = e_{[\mu p]}a' \in (A_{\psi=1}^\mu)^{[\mu p]}$$

so that $l(a) = l_2(a) = \mu p$ by Lemma 42. We have

$$\nabla a' \equiv \frac{\mu p}{e} \cdot \pi_k^{\mu p - e} + \frac{\mu p}{e} \cdot \pi_k^{\mu p} - \frac{\mu p + e(p-1)}{e} \cdot \pi_k^{\mu p - e + e(p-1)} + \ldots \pmod{p^2}$$

and hence

$$\nabla a = \nabla e_{[\mu p]} a' = e_{[\mu p - e]} \nabla a'$$

$$\equiv \frac{\mu p}{e} \cdot \pi_k^{\mu p - e} + \ldots \left( \frac{\mu p + e(p-1)}{e} - \frac{\mu p}{e} \right) \pi_k^{\mu p - e + e(p-1)} + \ldots \pmod{p^2},$$

where $x$ is the coefficient of $\pi_k^{\mu p - e + e(p-1)}$ in the expansion of $e_{[\mu p - e]} (\pi_k^{\mu p - e} + \pi_k^{\mu p})$. Moreover

$$l(\nabla a) = l(\nabla e_{[\mu p]} a') = l(e_{[\mu p - e]} \nabla a') = l(\nabla a') = \mu p - e + e(p-1).$$

In order to show that $v_\sigma(\nabla a) = l(\nabla a)$ write

$$\nabla a = \sum_i b_i \cdot \pi_K^i.$$
since then $p \mid b_i$. For $i < \mu p - e$ it suffices to show by (51) that we have instead

$$v_{\sigma}(a_i \sigma^i) > \mu p + e(p - 1)$$

for $i < \mu p$. Since $l_2(a) = \mu p$ we have $v_{\sigma}(a_i) \geq 2e(p - 1)$ for

$$\mu p - e(p - 1) < i < \mu p$$

and hence $v_{\sigma}(a_i \sigma^i) > \mu p + e(p - 1)$. For

$$\mu p - 2e(p - 1) < i \leq \mu p - e(p - 1)$$

we have by $v_{\sigma}(a_i) \geq 3e(p - 1)$ by Proposition 34(a) since

$$i \leq \mu p - e(p - 1) < -\left(\frac{3}{2} - \frac{2}{p}\right) \cdot e.$$ 

Indeed this last inequality is equivalent to

$$\mu p < \left( (p - 1) - \left(3 - \frac{2}{p}\right) \right) \cdot e \iff \mu < e - \left(\frac{4}{p} - \frac{2}{p^2}\right) \cdot e,$$

which holds by our assumption $4e < p$, noting that $e - 1$ is the maximal value for $\mu$.

Finally for

$$i \leq \mu p - 2e(p - 1) < -e - e(p - 1) = -ep$$

we have by (41)

$$v_{\sigma}(a_i \sigma^i) \geq -(p - 1)i - pe > (p - 1)pe - pe = (p - 2)pe$$

$$\geq (2p - 3)e = -e + 2e(p - 1) > \mu p + e(p - 1).$$

□

Proposition 51. If $e < p/4$ the identity (47) holds for $r = 2$.

Proof. By Lemmas 49 and 50 we can choose $\alpha_i$ such that

$$v_{\sigma}(\nabla \alpha_i) = l(\nabla \alpha_i) = \begin{cases} n_i - 2e & \text{if } p \nmid n_i \text{ and } p \nmid n_i - e, \\ n_i - 2e + e(p - 1) & \text{if } p \mid n_i \text{ or } p \mid n_i - e. \end{cases}$$

As in the proof of Proposition 47, for each $0 < n_1 < e$ there is a unique $n_2 = pn_1$ divisible by $p$. Similarly for each $n_h$ with $e < n_h < 2e$ (which is unique if it exists) there is a unique

$$n_{h+1} - e = p(n_h - e)$$

divisible by $p$. Note here that $n_h \leq 2e - 1$ and hence

$$n_{h+1} \leq p(e - 1) + e < e(p - 1)$$

using $2e < p$. Let

$$q_i := \nabla \alpha_i^{-1}\sqrt[p]{\xi_p - 1} \in K$$
be the basis of $L_2$. We again find that

\[
p \cdot q_1, p^{-1} \cdot q_2, \ldots, p \cdot q_h, p^{-1} \cdot q_{h+1}, \ldots, q_{f_n} \quad \text{if } n_1 < e \text{ and } e < n_h < 2e,
\]

\[
p \cdot q_1, p^{-1} \cdot q_2, \ldots, q_h, q_{h+1}, \ldots, q_{f_n} \quad \text{if } n_1 < e \text{ and } \not\exists \ e < n_h < 2e,
\]

\[
q_1, q_2, \ldots, p \cdot q_h, p^{-1} \cdot q_{h+1}, \ldots, q_{f_n} \quad \text{if } \not\exists n_1 < e \text{ and } e < n_h < 2e,
\]

\[
q_1, q_2, \ldots, q_h, q_{h+1}, \ldots, q_{f_n} \quad \text{if } \not\exists n_1 < e \text{ or } e < n_h < 2e,
\]

is a basis of $\mathcal{O}_K^{[n_1-2e]}$ and the statement follows. □

**Acknowledgements**

We would like to thank the referee for a very careful reading of the manuscript, which helped to improve our exposition a lot.

**References**


On the local Tamagawa number conjecture for Tate motives

Communicated by Kiran S. Kedlaya
Received 2015-08-25 Revised 2016-03-09 Accepted 2016-05-18

gdaigle@oxy.edu
Department of Mathematics, Occidental College, 1600 Campus Road, Los Angeles, CA 90041, United States

flach@caltech.edu
Department of Mathematics, Caltech, 253/37, Pasadena, CA 91125, United States


Heegner divisors in generalized Jacobians and traces of singular moduli

Jan Hendrik Bruinier and Yingkun Li

We prove an abstract modularity result for classes of Heegner divisors in the generalized Jacobian of a modular curve associated to a cuspidal modulus. Extending the Gross–Kohnen–Zagier theorem, we prove that the generating series of these classes is a weakly holomorphic modular form of weight \( \frac{3}{2} \). Moreover, we show that any harmonic Maass form of weight 0 defines a functional on the generalized Jacobian. Combining these results, we obtain a unifying framework and new proofs for the Gross–Kohnen–Zagier theorem and Zagier’s modularity of traces of singular moduli, together with new geometric interpretations of the traces with nonpositive index.

1. Introduction

The celebrated Gross–Kohnen–Zagier theorem [Gross et al. 1987] states that the generating series of Heegner divisors on the modular curve \( X_0(N) \) is a cusp form of weight \( \frac{3}{2} \) with values in the Jacobian of \( X_0(N) \). This result was later generalized by various authors to orthogonal and unitary Shimura varieties of higher dimension; see, e.g., [Borcherds 1999; Kudla 2004; Liu 2011].

In a different direction, Zagier [2002] proved that the traces of the normalized \( j \)-invariant over Heegner divisors of discriminant \( -d \) on the modular curve \( X(1) \) are the coefficients of a weakly holomorphic modular form of weight \( \frac{3}{2} \). This result was also generalized in subsequent work to modular curves of arbitrary level, traces of harmonic Maass forms over twisted Heegner divisors, and to cover more general nonpositive weight modular functions; see, e.g., [Alfes and Ehlen 2013; Bringmann et al. 2005; Bruinier and Funke 2006; Duke and Jenkins 2008; Funke 2002; Kim 2004]. Recently, Gross [2012] has explained how Zagier’s original result can be related to their earlier joint result with Kohnen. He showed that the traces of singular moduli on \( X(1) \) can be interpreted in terms of Heegner divisors in the generalized Jacobian associated with the modulus \( 2 \cdot (\infty) \).

The authors are partially supported by DFG grant BR-2163/4-1.


Keywords: Singular moduli, generalized Jacobian, Heegner point, Borcherds product, harmonic Maass form.
We pick up this idea of Gross and define classes of Heegner divisors of arbitrary discriminant in the generalized Jacobian $J_m(X)$ of a modular curve $X$ of arbitrary level with cuspidal modulus $m$. Then we prove that the generating series of these classes is a weakly holomorphic modular form of weight $\frac{3}{2}$ with values in $J_m(X)$. Our argument is a generalization of Borcherds’ proof [1999] of the Gross–Kohnen–Zagier theorem [Borcherds 1999] and relies on the construction of explicit relations among Heegner divisors given by automorphic products. Note that, in contrast to [Borcherds 1999], we need to use the explicit infinite product expansions of automorphic products at all cusps of $X$. By applying the natural map between $J_m(X)$ and the usual Jacobian $J(X)$ to this generating series we recover the “classical” Gross–Kohnen–Zagier theorem.

Then we show that every harmonic Maass form $F$ of weight 0 on $X$ with vanishing constant term at every cusp (such as the normalized $j$-function when $X = X(1)$) defines a functional $\text{tr}_F$ on $J_m(X)$. The value of $\text{tr}_F$ on Heegner divisors of negative discriminant $-d$ is just the sum of the values of $F$ over the Heegner points of discriminant $-d$. The value of $\text{tr}_F$ on “Heegner divisors” of nonnegative discriminant can be explicitly computed in terms of the principal parts of $F$ at the cusps. In that way we are able to recover Zagier’s result and its generalizations in [Alfes and Ehlen 2013; Bruinier and Funke 2006].

We now describe the content of the present paper in more detail. To simplify the exposition, throughout this introduction we let $p$ be prime or 1 and consider the modular curve $X^*_0(p)$ associated to the extension $0^*_0(p)$ of $0^*_0(p)$ in $\text{PSL}_2(\mathbb{Z})$ by the Fricke involution. In the body of this paper, we consider modular curves of arbitrary level (as modular curves associated to orthogonal groups of signature $(1, 2)$).

Let $\infty$ be the cusp of $X^*_0(p)$ and let $m$ be a nonnegative integer. Then $m = m \cdot (\infty)$ is an effective divisor. Recall that the generalized Jacobian $J_m(X^*_0(p))$ of $X^*_0(p)$ associated with the modulus $m$ is a commutative algebraic group whose rational points correspond to classes of divisors of degree zero modulo $m$-equivalence; see Section 2 and [Serre 1988]. If $m = 0$, then $J_m(X^*_0(p))$ is simply the usual Jacobian.

For any integer $d$, let $Q_{p,d}$ be the set of (positive definite if $d > 0$) integral binary quadratic forms $[a, b, c]$ of discriminant $-d = b^2 - 4ac$ with $p$ dividing $a$. If $d \neq 0$ then $\Gamma^*_0(p)$ acts on $Q_{p,d}$ with finitely many orbits.

If $d$ is positive, then any $Q \in Q_{p,d}$ defines a point $\alpha_Q$ in the upper complex half-plane $\mathbb{H}$, the solution of the equation $az^2 + bz + c = 0$ with positive imaginary part. There is a corresponding Heegner divisor of discriminant $-d$ on $X^*_0(p)$ given by

$$Y(d) = \sum_{Q \in Q_{p,d} / \Gamma^*_0(p)} \frac{1}{|\Gamma^*_0(p)_Q|} \cdot (\alpha_Q),$$

where $\Gamma^*_0(p)_Q$ is the (finite) stabilizer of $Q$ (see Equation (1.5) in [Bruinier and...\]
Funke 2006]. The divisor

\[ Z(d) = Y(d) - \text{deg}(Y(d)) \cdot (\infty) \]

has degree zero and is defined over \( \mathbb{Q} \). We denote by \([Z(d)]_m\) its class in the generalized Jacobian \( J_m(X^*_0(p)) \).

If \( d \) is negative, any \( Q \in \mathbb{Q}_{p,d} \) defines an oriented geodesic cycle on \( \mathbb{H} \cup P^1(\mathbb{R}) \), given by the equation \( a|z|^2 + b \Im(z) + c = 0 \). It has nontrivial intersection with \( P^1(\mathbb{Q}) \) if and only if \( d \) is the negative of a square of an integer. In this case the two solutions in \( P^1(\mathbb{Q}) \) define cusps of the modular curve. There is a unique cusp \( c_Q \in P^1(\mathbb{Q}) \) from which the geodesic originates. (In the present \( \Gamma_0^*(p) \) example all cusps collapse to \( \infty \) under the map to the quotient, but this is of course not true for more general congruence subgroups.) If \( d = -b^2 \) for a nonzero integer \( b \), then \( Q \) is \( \Gamma_0^*(p) \)-equivalent to \([0, b, c]\) with \( c \in \mathbb{Z}/b\mathbb{Z} \) and \( c_Q \) is equivalent to \( \infty \). We let \( h_Q \in \mathbb{Q}(X^*_0(p))^\times \) be a function satisfying

\[ h_Q = 1 - q^b_\infty + O(q^m_\infty) \]

at the cusp \( \infty \), where \( q_\infty \) is the uniformizing parameter of the completed local ring at \( \infty \) given by the Tate curve over \( \mathbb{Z}[q_\infty] \). Then we define

\[ [Z(d)]_m = [\operatorname{div}(h_{[0,b,0]})]_m = \sum_{Q \in \mathbb{Q}_{p,d}/\Gamma_0^*(p)} \frac{1}{b} \cdot [\operatorname{div}(h_Q)]_m. \]

Note that this class vanishes if \( d \leq -m^2 \). If \( d < 0 \) is not the negative of the square of an integer, we put \([Z(d)]_m = 0 \). Finally, for \( d = 0 \), we define \([Z(0)]_m \) as the class of the line bundle of modular forms \( \mathcal{M}_{-1} \) of weight \(-1\) on \( X^*_0(p) \) (see Section 2 for details).

To describe the relations among the classes \([Z(d)]_m\), we consider the generating series

\[ A_m(\tau) = \sum_{d \in \mathbb{Z}} [Z(d)]_m \cdot q^d \in \mathbb{C}((q)) \otimes J_m(X^*_0(p)). \]

It is a formal Laurent series in the variable \( q = e^{2\pi i \tau} \) for \( \tau \in \mathbb{H} \). Our first main result is the following (see also Theorem 4.2).

**Theorem 1.1.** The generating series \( A_m(\tau) \) is a weakly holomorphic modular form of weight \( \frac{3}{2} \) for the group \( \Gamma_0(4p) \), that is, \( A_m(\tau) \in M^!_{3/2}(\Gamma_0(4p)) \otimes J_m(X^*_0(p)). \)

Under the natural map

\[ J_m(X^*_0(p)) \rightarrow J(X^*_0(p)) \]

the classes \([Z(d)]_m\) with \( d \leq 0 \) are mapped to zero. Applying it to \( A_m(\tau) \), we recover the Gross–Kohnen–Zagier theorem (see also Corollary 4.5).
Corollary 1.2 (Gross–Kohnen–Zagier). The generating series $A_0(\tau)$ of classes of Heegner divisors $[Z(d)]_0$ in the Jacobian is a cusp form of weight $\frac{3}{2}$ for the group $\Gamma_0(4p)$, that is, $A_0(\tau) \in S_{3/2}(\Gamma_0(4p)) \otimes J(X_0^*(p))$.

To recover the results of [Zagier 2002] and [Bruinier and Funke 2006] on traces of modular functions from Theorem 1.1, we show that harmonic Maass forms define functionals on $J_m(X_0^*(p))$. Let $F \in H_0^+((\Gamma_0^*)^*(p))$ be a harmonic Maass form for $\Gamma_0^*(p)$ of weight 0 as in [Bruinier and Funke 2004]. Denote the Fourier expansion of the holomorphic part of $F$ by

$$F^+(\tau) = \sum_{n \gg -\infty} c_F^+(n) \cdot q^n.$$

Proposition 1.3. Assume that $c_F^+(n) = 0$ for $n \leq -m$ and $c_F^+(0) = 0$. Then there is a linear map $\text{tr}_F : J_m(X_0^*(p)) \to \mathbb{C}$ defined by

$$[D]_m \mapsto \text{tr}_F(D) := \sum_{a \in \text{supp}(D) \setminus \{\infty\}} n_a \cdot F(a),$$

for divisors $D = \sum_a n_a \cdot (a)$ in $\text{Div}_0(X_0^*(p))$.

The images under $\text{tr}_F$ of the classes $[Z(d)]_m$ with $d \leq 0$ can be explicitly computed in terms of the principal part of $F$. As a consequence we derive the following theorem (see also Theorem 5.2).

Theorem 1.4. The series $\text{tr}_F(A_m)$ is a weakly holomorphic modular form in the space $M_{3/2}^!(\Gamma_0(4p))$. It is explicitly given by

$$\text{tr}_F(A_m) = \sum_{d > 0} F(Y(d)) \cdot q^d + \sum_{n \geq 1} c_F^+(-n)(\sigma_1(n) + p\sigma_1(n/p))$$

$$- \sum_{b > 0} \sum_{n > 0} c_F^+(-bn) \cdot b \cdot q^{-b^2}.$$

The modularity of the right-hand side was also proved in [Bruinier and Funke 2006] by interpreting it as the Kudla–Millson theta lift of $F$. Applying this theorem to the special case where $p = 1$, $m \geq 2$, and $F = j - 744$, Zagier’s original result on traces of singular moduli can be obtained.

In the body of the paper we work with modular curves of arbitrary level associated with orthogonal groups of even lattices of signature $(1,2)$. This setup is natural, since the proof of Theorem 1.1 implicitly relies on the singular theta correspondence for the dual reductive pair given by $\text{SL}_2$ and $\text{O}(1,2)$. For the modulus we allow arbitrary effective divisors that are supported on the cusps. The generating series of Heegner divisors is then a vector-valued modular form for the metaplectic extension of $\text{SL}_2(\mathbb{Z})$ transforming with the Weil representation of a finite quadratic module.
In Section 2 we recall some basic facts on generalized Jacobians of curves. Section 3 contains our setup for modular curves associated to orthogonal groups, Heegner divisors, and vector-valued modular forms. Then we define classes of Heegner divisors in generalized Jacobians in Section 4, and prove the abstract modularity theorem for these classes. In Section 5 we prove that harmonic Maass forms define functionals on the generalized Jacobian and derive modularity results for the traces of harmonic Maass forms over Heegner divisors from the abstract modularity theorem. We also give some explicit examples and indicate possible generalizations in Section 6.

2. Generalized Jacobians

Let \( X \) be a complete nonsingular algebraic curve over a field \( k \) of characteristic 0. Let \( \text{Div}^0(X) \) be the group of divisors of \( X \) of degree 0 defined over \( k \), and denote by \( P(X) \) the subgroup of divisors of rational functions \( f \in k(X)^\times \). The Jacobian \( J(X) \) of \( X \) is a commutative algebraic group over \( k \) whose \( k \)-rational points are isomorphic to the quotient group \( \text{Div}^0(X)/P(X) \).

Recall that there is the notion of the generalized Jacobian; see, e.g., [Serre 1988, Chapter 5] for details. Let \( S \subset X(k) \) be a finite set of points, and for \( s \in S \) let \( m_s \in \mathbb{Z}_{\geq 0} \). Then

\[
m = \sum_{s \in S} m_s \cdot (s)
\]

is an effective divisor defined over \( k \). Let \( O_s \) be the ring of integers in the completion \( k(X)_s \) of \( k(X) \) at \( s \), and let \( \pi_s \in O_s \) be a uniformizer. If \( f, g \in k(X)_s \) and \( n \in \mathbb{Z} \), we write

\[
f = g + O(\pi_s^n)
\]

if \( f - g \in \pi_s^n O_s \). We consider the subgroup

\[
P_m(X) = \{ \text{div}(f) : f \in k(X)^\times \text{ with } \pi_s^{-\text{ord}_s(f)} f = 1 + O(\pi_s^{m_s}) \text{ for all } s \in S \}
\]

of \( P(X) \). The generalized Jacobian \( J_m(X) \) associated with the modulus \( m \) is a commutative algebraic group over \( k \), whose \( k \)-rational points satisfy

\[
J_m(X)(k) \cong \text{Div}^0(X)/P_m(X).
\] (2-1)

The quotient on the right hand side is also canonically isomorphic to the subgroup of divisors in \( \text{Div}^0(X) \) coprime to \( S \) modulo \( m \)-equivalence. For a divisor \( D \in \text{Div}^0(X) \) we denote by \([D]_m \) the corresponding class in \( J_m(X)(k) \).

There is a canonical rational map \( \varphi_m : X \to J_m(X) \) defined over \( k \) which is regular outside \( S \), see [Serre 1988, Chapter 5, Theorem 1]. If \( m' \) is another effective divisor on \( X \) satisfying \( m \geq m' \geq 0 \), there exists a unique homomorphism \( J_m \to J_{m'} \)
which is compatible with \( \varphi_m \) and \( \varphi_{m'} \). It is surjective and separable [Serre 1988, Chapter 5, Proposition 6]. In particular, there exists a surjective homomorphism

\[
J_m(X) \to J(X).
\]

(2-2)

Its kernel is isomorphic to

\[
H_m = \left( \prod_{s \in S} \mathbb{G}_m \times \mathbb{G}_m^{m_s-1} \right) / \mathbb{G}_m,
\]

(2-3)

where the quotient is with respect to the diagonally embedded multiplicative group. Typical elements of the kernel are obtained, by choosing a pair \((s, n)\) with \(s \in S\) and \(n > 0\) and a function \(h_{s,n} \in k(X)^\times\) such that

\[
h_{s,n} = \begin{cases} 
1 - \pi^n_s + O(\pi^{m_s}_s) & \text{at } s, \\
1 + O(\pi^{m'_t}_t) & \text{at all } t \in S \setminus \{s\}.
\end{cases}
\]

(2-4)

An argument as in [Serre 1988, Chapter 5, Proposition 8] shows that the “additive part” of \( H_m \) is generated by the classes

\[
[\text{div}(h_{s,n})]_m,
\]

(2-5)

for \(s \in S\) and \(0 < n < m_s\). Note that for \(n \geq m_s\) the class \([\text{div}(h_{s,n})]_m\) vanishes.

Let \(s_0 \in S\) be a fixed base point. If \(\mathcal{L}\) is a line bundle on \(X\) which is defined over \(k\), and \((\phi_s)_{s \in S}\) is a family of local trivializations of \(\mathcal{L}\) at the points of \(S\), we can associate to the pair \((\mathcal{L}, (\phi_s))\) a class in \(J_m(X)\) as follows. It is easily seen that there exists a rational section \(f\) of \(\mathcal{L}\) such that

\[
\phi_s^{-1} f = \pi_s^{a_s} \cdot (1 + O(\pi^{m_s}_s)),
\]

(2-6)

for some \(a_s \in \mathbb{Z}\) at every \(s \in S\). Then we define

\[
[(\mathcal{L}, (\phi_s))]_m = [\text{div}(f) - \text{deg}(\mathcal{L}) \cdot (s_0)]_m \in J_m(X)(k).
\]

(2-7)

3. Modular curves

Here we recall the description of modular curves as Shimura varieties associated to orthogonal groups. We also define classes of Heegner divisors in generalized Jacobians.

Let \((L, Q)\) be an isotropic even lattice of signature \((1, 2)\). We denote by \((x, y)\) the bilinear form corresponding to the quadratic form \(Q\), normalized such that \(Q(x) = \frac{1}{2}(x, x)\). For any commutative ring \(R\) we write \(L_R = L \otimes_{\mathbb{Z}} R\). Throughout we fix an orientation on \(L_{\mathbb{R}}\), and write \(L'\) for the dual lattice of \(L\). Let

\[
N = \min\{n \in \mathbb{Z}_{>0} : nQ(\lambda) \in \mathbb{Z} \text{ for all } \lambda \in L'\}
\]
be the level of $L$, and denote by $\text{disc}(L) = |L'/L|$ the discriminant of $L$. We let $\text{SO}(L)$ be the special orthogonal group of $L$ and write $\text{SO}^+(L)$ for the intersection of $\text{SO}(L)$ with the connected component of the identity of $\text{SO}(L)(\mathbb{R})$. The even Clifford algebra of $L_{\mathbb{Q}}$ is isomorphic to the matrix algebra $\text{Mat}_2(\mathbb{Q})$, which induces an isomorphism $\text{PGL}_2(\mathbb{Q}) \cong \text{SO}(L)(\mathbb{Q})$. We realize the hermitian symmetric space corresponding to $\text{SO}(L)$ as the domain

$$D = \{z \in L_{\mathbb{C}} : (z, z) = 0, (z, \bar{z}) < 0\}/\mathbb{C}^\times.$$ 

It decomposes into 2 connected components. We fix one of these components and denote it by $D^+$. Let $\Gamma = \Gamma_L$ be the discriminant kernel subgroup of $\text{SO}^+(L)$, that is, the kernel of the natural homomorphism

$$\text{SO}^+(L) \to \text{Aut}(L'/L).$$

Recall that rescaling the quadratic form by a factor of $n$ does not change $\text{SO}^+(L)$ while it replaces the discriminant kernel by the full congruence subgroup of level $n$. We denote by

$$Y_{\Gamma} = \Gamma \backslash D^+$$

the noncompact modular curve associated with $\Gamma$.

Let $\text{Iso}(L)$ be the set of isotropic lines in $L$ (i.e., primitive isotropic rank-1 sublattices $I \subset L$). The group $\Gamma$ acts with finitely many orbits on $\text{Iso}(L)$. We denote by $X_{\Gamma}$ the compact modular curve obtained by adding to $Y_{\Gamma}$ the cusps corresponding to the $\Gamma$-classes of isotropic lines $I \in \text{Iso}(L)$; see, e.g., [Bruinier and Funke 2006]. It is well known that $X_{\Gamma}$ is a projective algebraic curve which has a canonical model over a cyclotomic field.

As in [Bruinier and Funke 2006], we choose an orientation on the isotropic lines as follows. We fix one line $I_0 \in \text{Iso}(L)$ together with an orientation on $I_0$ given by a basis vector $x_0 \in I_{0,\mathbb{R}}$. For any other $I \in \text{Iso}(L)$ we choose a $g \in \text{SO}^+(L)(\mathbb{R})$ such that $g I_{0,\mathbb{R}} = I_{\mathbb{R}}$. Then $g x_0 \in I_{\mathbb{R}}$ defines an orientation on $I$ which is independent of the choices of $g$ and $x_0$.

Let $I \subset L$ be a primitive isotropic line and write $c_I \in X_{\Gamma}$ for the cusp corresponding to $I$. Local coordinates near $c_I$ can be described as follows. We write $N_I$ for the positive generator of the ideal $(I, L) \subset \mathbb{Z}$. It is a divisor of $N$. Throughout, we let $\ell = \ell_I$ be the positive generator of $I$ and fix a vector $\ell' = \ell'_I \in L'$ such that

$$(\ell, \ell') = 1.$$ 

We let $K$ be the even negative definite lattice

$$K = L \cap \ell^\perp \cap \ell'^\perp.$$ 

(3-2)

(3-3)
If $\ell_K \in K$ denotes a generator, then $K$ is isomorphic to $\mathbb{Z}$ equipped with the quadratic form $x \mapsto Q(\ell_K)x^2$. The quantity $4Q(\ell_K)$ divides $N$. The holomorphic map

$$\mathbb{H} \to \mathcal{D}, \quad w \mapsto \mathbb{C}^\times(w \otimes \ell_K + \ell' - Q(w \otimes \ell_K)\ell - Q(\ell')\ell) \quad (3-4)$$

is injective and has one of the two connected components of $\mathcal{D}$ as its image. Possibly replacing $\ell_K$ by its negative, we may assume that this map is an isomorphism from $\mathbb{H}$ onto $\mathcal{D}^+$. It is compatible with the natural actions of $\text{PGL}_2^+(\mathbb{Q})$ on $\mathbb{H}$ by fractional linear transformations and on $\mathcal{D}^+$ via the isomorphism with $\text{SO}^+(L)(\mathbb{Q})$. For $\mu \in L_\mathbb{Q} \cap I^\perp$ we consider the Eichler transformation

$$E_{\ell,\mu}(x) = x + (x, \ell)\mu - (x, \mu)\ell - (x, \ell)Q(\mu)\ell \quad (3-5)$$

in $\text{SO}^+(L)(\mathbb{Q})$. It belongs to $\Gamma$ if $\mu \in K$.

**Lemma 3.1.** The stabilizer in $\Gamma$ of the primitive isotropic line $I$ is given by

$$\Gamma_I = \{E_{\ell,\mu} : \mu \in K\}.$$ 

**Proof.** Let $\gamma \in \Gamma_I$. Then $\gamma\ell = \pm\ell$. We first assume that $\gamma\ell = \ell$. Then

$$u := \gamma\ell' - \ell'$$

belongs to $L \cap \ell^\perp$, and $v := u - (u, \ell')\ell$ belongs to $K$. It is easily checked that

$$E_{\ell,v}(\ell) = \ell, \quad E_{\ell,v}(\ell') = \gamma\ell'.$$

Hence $\gamma^{-1}E_{\ell,v}$ leaves the vectors $\ell$ and $\ell'$ fixed. Consequently, it maps the orthogonal complement $K$ to itself, and therefore $\ell_K$ to $\pm\ell_K$. Since $\gamma^{-1}E_{\ell,v}$ has determinant 1, the sign must be positive and thus $\gamma = E_{\ell,v}$.

We now consider the case $\gamma\ell = -\ell$. The orthogonal transformation $\sigma$ taking $\ell$ to $-\ell$, and $\ell'$ to $-\ell'$, and $\ell_K$ to itself belongs to $\text{SO}(L)(\mathbb{Q})$. The element $\sigma\gamma \in \text{SO}(L)(\mathbb{Q})$ fixes $\ell$. Arguing as above, we see that it is equal to an Eichler transformation $E_{\ell,u} \in \text{SO}^+(L)(\mathbb{Q})$. This implies that $\sigma$ belongs to the connected component of the identity of $\text{SO}(L)(\mathbb{R})$. But this leads to a contradiction, since the spinor norm of $\sigma$ is negative, showing that the case $\gamma\ell = -\ell$ cannot occur. \qed

The action of $\mathbb{Z}$ on $\mathbb{H}$ by translations corresponds to the action of $\Gamma_I$ on $\mathcal{D}^+$. The induced map

$$\mathbb{Z}\backslash \mathbb{H} \to \Gamma_I \backslash \mathcal{D}^+ \quad (3-6)$$

is an isomorphism. Hence, $q_I = e^{2\pi i w}$ defines a local parameter at the cusp $c_I$ of $X_\Gamma$.

**Example 3.2.** In the special case when $N_I = 1$, then $4N = -Q(\ell_K)$ and the discriminant kernel subgroup $\Gamma$ is isomorphic to $\Gamma_0(N/4)$. The curve $X_\Gamma$ is isomorphic to $X_0(N/4)$, with $c_I$ corresponding to the cusp at $\infty$; see, e.g., [Bruinier and Ono 2010, Section 2.4].
The Weil representation. Let $\text{Mp}_2(\mathbb{Z})$ be the metaplectic extension of $\text{SL}_2(\mathbb{Z})$ by \{±1\}, realized by the two possible choices of a holomorphic square root of the automorphy factor $c\tau + d$ for $(a\ b\ c\ d) \in \text{SL}_2(\mathbb{Z})$; see, e.g., [Borcherds 1998; Kudla 2003].

Recall that there is a Weil representation $\omega_L$ of $\text{Mp}_2(\mathbb{Z})$ on the complex vector space $S_L$ of functions $L'/L \to \mathbb{C}$ on the discriminant group. Identifying $S_L$ with the space of Schwartz–Bruhat functions on $L \otimes \hat{\mathbb{Q}}$ which are supported on $L' \otimes \hat{\mathbb{Z}}$ and translation invariant under $L \otimes \hat{\mathbb{Z}}$, the representation $\omega_L$ can be viewed as the restriction of the usual Weil representation of $\text{Mp}_2(\hat{\mathbb{Q}})$ on $L \otimes \hat{\mathbb{Q}}$ with respect to the standard additive character of $\hat{\mathbb{Q}}$; see [Kudla 2003]. The representation $\omega_L$ is the complex conjugate of the representation $\rho_L$ in [Borcherds 1998; Bruinier 2002; Bruinier and Funke 2006]. The action of $\text{Mp}_2(\mathbb{Z})$ on $S_L$ commutes with the natural action of $\text{Aut}(L'/L)$ by translation of the argument.

If $k \in \frac{1}{2}\mathbb{Z}$, we denote by $M^!_k(\omega_L)$ the space of $S_L$-valued weakly holomorphic modular forms for $\text{Mp}_2(\mathbb{Z})$ of weight $k$ with representation $\omega_L$. The subspace of holomorphic modular forms is denoted by $M_k(\omega_L)$.

Heegner divisors. For any $d \in \mathbb{Q}^\times$, the group $\Gamma$ acts on the set
\[ L'_d = \{ \lambda \in L' : Q(\lambda) = d \} \]
with finitely many orbits. For every $\lambda \in L'$ with $Q(\lambda) > 0$, the stabilizer $\Gamma_\lambda \subset \Gamma$ of $\lambda$ is finite, and there is a unique point $z_\lambda \in D^+$ which is orthogonal to $\lambda$. For $d \in \mathbb{Q}_{>0}$ and $\varphi \in S_L$ we consider the Heegner divisor
\[ Y(d, \varphi) = \sum_{\lambda \in L'_d/\Gamma} \frac{1}{2|\Gamma_\lambda|} \varphi(\lambda) \cdot (z_\lambda) \quad (3-7) \]
on $X_\Gamma$. It is defined over the field of definition of $X_\Gamma$ and has coefficients in the field of definition of $\varphi$. Let $I_0 \in \text{Iso}(L)$ be a fixed isotropic line. We define a divisor of degree 0 on $X_\Gamma$ by putting
\[ Z(d, \varphi) = Y(d, \varphi) - \text{deg}(Y(d, \varphi)) \cdot (c_{I_0}). \quad (3-8) \]


We now consider classes of Heegner divisors in the generalized Jacobian of the modular curve $X := X_\Gamma$ as defined in the previous section. We let $k \subset \mathbb{C}$ be the number field obtained by adjoining the primitive root of unity $e^{2\pi i/N}$ to the common field of definition of the canonical model and all of the cusps of $X$. Let $S = \{ c_I : I \in \text{Iso}(L)/\Gamma \}$ be the set of cusps of $X$ and let
\[ m = \sum_{I \in \text{Iso}(L)/\Gamma} m_I \cdot (c_I) \]
be a fixed effective divisor supported on $S$. We consider the generalized Jacobian of $X$ associated with the modulus $m$. For $I \in \text{Iso}(L)$, we take as the uniformizing parameter in the completed local ring at $c_I$ the parameter $q_I = e^{2\pi i w}$ defined by (3-6) (given by the Tate curve over $\mathbb{Z}[\|q_I\|]$ when $N_I = 1$ such that $X_I \cong X_0(N/4)$).

Since, throughout this section, we are only interested in the $k$-valued points of the generalized Jacobian, we briefly write $J_m(X)$ instead of $J_m(X)(k)$. For every degree-zero divisor $D = \sum a_I \cdot (c_I) \in \text{Div}^0(X)$ supported on $S$ and every tuple $r = (r_I) \in \mathbb{G}_m^{[S]}(k)$, we choose a function $u_{D,r} \in k(X)^\times$ such that

$$u_{D,r} = r_I q_I^{a_I} \cdot (1 + O(q_I^m))$$

(4-1)

at $c_I$ for $I \in \text{Iso}(L)$. We write $H_{\mathbb{G}_m,m}$ for the subgroup of $J_m(X)$ generated by the classes $[\text{div}(u_{D,r})]_m$ of all these functions and let

$$J_m^{\text{add}}(X) = J_m(X)/H_{\mathbb{G}_m,m}.$$  

(4-2)

By definition we have $J_m^{\text{add}}(X) = J_m(X)$ when $|S| = 1$. By the Manin–Drinfeld theorem we have $J_m^{\text{add}}(X)_Q = J_m(X)_Q$ when $m = 0$. For general $m$ the kernel of the induced homomorphism

$$J_m^{\text{add}}(X)_Q \to J(X)_Q$$

(4-3)

is isomorphic to the product of the groups $\mathbb{G}_m^{m_I-1}$ for $I \in \text{Iso}(L)/\Gamma$ with $m_I > 0$.

For $d \in \mathbb{Q}_{>0}$ and $\varphi \in S_L$ we consider the class

$$[Z(d, \varphi)]_m \in J_m(X)_C$$

(4-4)

of the Heegner divisor $Z(d, \varphi)$ in the generalized Jacobian.

Let $\mathcal{T}$ be the tautological bundle on $X$, and define the line bundle of modular forms of weight $2k$ on $X$ by $\mathcal{M}_{2k} = \mathcal{T}^{\otimes k}$. (Sections of $\mathcal{M}_{2k}$ correspond to classical elliptic modular forms of weight $2k$ under the isomorphism $\text{SO}(L)(\mathbb{Q}) \cong \text{PGL}_2(\mathbb{Q})$.) Recall that $\mathcal{T}$ is canonically trivial in small neighborhoods of the cusps. Hence, taking the induced trivializations and putting $s_0 = c_{t_0}$ in (2-7), we obtain a class $[\mathcal{M}_k]_m \in J_m(X)_Q$. For $d = 0$ we define

$$[Z(0, \varphi)]_m = \varphi(0) \cdot [\mathcal{M}_{-1}]_m.$$  

(4-5)

We also define classes for $d \in \mathbb{Q}_{<0}$ as follows. For a vector $\lambda \in L'_{\text{d}}$, the orthogonal complement $\lambda^\perp \subset L_Q$ is isotropic if and only if $d = -2 \text{disc}(L)(\mathbb{Q}^\times)^2$. In this case there is a unique pair of isotropic lines $I, \tilde{I} \in \text{Iso}(L)$ such that $\lambda^\perp = I_Q + \tilde{I}_Q$ and such that the triple $(\lambda, x, \tilde{x})$ is a positively oriented basis of $L_Q$ for positive basis vectors $x \in I$ and $\tilde{x} \in \tilde{I}$. Following [Bruinier and Funke 2006], we call $I$ the isotropic line associated to $\lambda$ and write $I \sim \lambda$. Note that $\tilde{I}$ is the isotropic line associated to $-\lambda$. We define the $I$-content $n_I(\mu)$ of any $\mu \in L' \cap I^\perp$ as follows: If $Q(\mu) = 0$ we put
\( n_I(\mu) = 0 \). If \( Q(\mu) \neq 0 \) we let \( n_I(\mu) \) be the unique nonzero integer such that
\[
(\mu, L \cap I^\perp) = n_I(\mu) \cdot \mathbb{Z}
\] (4-6)
and \( \text{sgn}(n_I(\mu)) \cdot \mu \sim I \).

Now, if \( d \in -2 \text{disc}(L)(\mathbb{Q}^\times)^2 \) and \( \lambda \in L_d' \), we let \( I \in \text{Iso}(L) \) be the isotropic line associated to \( \lambda \) and let \( \ell' \in \mathbb{L}' \) be as in (3-2) such that \((\ell', I) = \mathbb{Z} \). We choose a function \( h_\lambda \in k(X) \times \) such that
\[
1 - e^{2\pi i (\lambda, L')} q^n(\lambda) + O(q^m) \text{ at the cusp } c_I,
\]
\[
1 + O(q^m) \text{ at all other cusps } c_J.
\] (4-7)
The existence of \( h_\lambda \) follows for instance from the approximation theorem for valuations, see page 29 in [Serre 1988]. If \((\lambda, L') \in \mathbb{Z} \), then \( h_\lambda \) agrees with the function \( h_{cI,nI(\lambda)} \in k(X) \times \) defined in (2-4). For \( \varphi \in S_L \) we define
\[
[Z(d, \varphi)]_m = \sum_{\lambda \in L_d' / \Gamma} \frac{1}{2n_I(\lambda)} (\varphi(\lambda) + \varphi(-\lambda)) \cdot [\text{div}(h_\lambda^{-1})]_m.
\] (4-8)

It is easily checked that this class is independent of the choices of the functions \( h_\lambda \).

If \( d < 0 \) and \( d \notin -2 \text{disc}(L)(\mathbb{Q}^\times)^2 \), we put \([Z(d, \varphi)]_m = 0 \).

Finally, for all \( d \in \mathbb{Q} \) we write \([Z(d)]_m \) for the element of
\[
\text{Hom}(S_L, J_m(X)_C) \cong J_m(X)_C \otimes S_L^\vee
\]
given by \( \varphi \mapsto [Z(d, \varphi)]_m \).

The classes \([Z(d, \varphi)]_m \) with \( d < 0 \) can also be expressed in a slightly different way. To this end, for an isotropic line \( I \) we define
\[
L_{d,I}' = \{ \lambda \in L_d' : \lambda \perp I \text{ and } \lambda \sim I \}.
\]

**Lemma 4.1.** For \( d < 0 \) we have
\[
[Z(d, \varphi)]_m = \sum_{I \in \text{Iso}(L) / \Gamma} \sum_{\lambda \in L_{d,I}' / I} \frac{1}{2} (\varphi(\lambda) + \varphi(-\lambda)) \cdot [\text{div}(h_\lambda^{-1})]_m.
\]

**Proof.** If \( L_{d,I} \) is nonempty and if we fix \( \lambda_0 \in L_{d,I}' \), we have
\[
L_{d,I}' = \{ \lambda_0 + a \ell / N_I : a \in \mathbb{Z} \},
\]
\[
L_{d,I}' / \Gamma I = \{ \lambda_0 + a \ell / N_I : a \in \mathbb{Z} / N_I n_I(\lambda_0) \mathbb{Z} \},
\]
\[
L_{d,I}' / I = \{ \lambda_0 + a \ell / N_I : a \in \mathbb{Z} / N_I \mathbb{Z} \}.
\]

This implies the assertion. \( \square \)
An abstract modularity theorem. To describe the relations in the generalized Jacobian among the classes $[Z(d)]_m$ we form the generating series

$$A_m(\tau) = \sum_{d \in \frac{1}{N}\mathbb{Z}} [Z(d)]_m \cdot q^d \in S_L^\vee((q)) \otimes J_m^\text{add}(X)_C.$$  \hspace{1cm} (4-9)

It is a formal Laurent series in the variable\(^1\) $q = e^{2\pi i \tau}$, where $\tau \in \mathbb{H}$, with exponents in $\frac{1}{N}\mathbb{Z}$ and coefficients in $S_L^\vee \otimes J_m^\text{add}(X)_C$.

**Theorem 4.2.** The generating series $A_m(\tau)$ is the $q$-expansion of a weakly holomorphic modular form in $M_{3/2}(\omega^\vee_L) \otimes J_m^\text{add}(X)_C$.

To prove this result, we use the following variant of Borcherds’ modularity criterion [Borcherds 1999, Theorem 3.1]. Let $\rho$ be a finite dimensional representation of $\text{Mp}_2(\mathbb{Z})$ on a complex vector space $V$ which is trivial on some congruence subgroup. The stabilizer in $\text{Mp}_2(\mathbb{Z})$ of the cusp $\infty$ is generated by the elements $T = \left( \begin{array}{cc} 1 & 1 \\ 0 & 1 \end{array} \right)$ and $Z = \left( \begin{array}{cc} -1 & 0 \\ 0 & -1 \end{array} \right), i$. The hypothesis on $\rho$ implies that some power of $\rho(T)$ is the identity, and therefore all eigenvalues of $\rho(T)$ are roots of unity.

If $g \in M^k(\rho)$ is a weakly holomorphic modular form for $\text{Mp}_2(\mathbb{Z})$ of weight $k \in \frac{1}{2}\mathbb{Z}$ with representation $\rho$, then it has a Fourier expansion

$$g(\tau) = \sum_{n \in \mathbb{Q}} a(n) \cdot q^n,$$

where the coefficients $a(n) \in V$ satisfy the conditions

$$\rho(T)a(n) = e^{2\pi in}a(n),$$  \hspace{1cm} (4-10)

$$\rho(Z)a(n) = e^{-\pi ik}a(n).$$  \hspace{1cm} (4-11)

We write $\rho^\vee$ for the representation dual to $\rho$, and denote by $(\cdot, \cdot)$ the natural pairing $V \times V^\vee \to \mathbb{C}$.

**Proposition 4.3.** A formal Laurent series

$$g(\tau) = \sum_{n \in \mathbb{Q}} a(n) \cdot q^n \in V((q)),$$

with coefficients $a(n)$ satisfying the conditions (4-10) and (4-11) is the $q$-expansion of a weakly holomorphic modular form in $M^k(\rho)$ if and only if

$$\sum_{n \in \mathbb{Q}} (a(n), c(-n)) = 0$$

for all

$$f(\tau) = \sum_{n \in \mathbb{Q}} c(n) \cdot q^n \in M^1_{2-k}(\rho^\vee).$$

---

\(^1\)Confusion with the local parameter $q_I$ at the cusp $c_I$ of $X$ should not be possible.
Proof. This result is proved in Section 3 of [Borcherds 1999] in the special case when $g$ is actually a formal power series. The same proof applies to our slightly more general case, if we replace the vector bundle of modular forms of type $\rho$ by a twist with a power of the line bundle $L(\infty)$ corresponding to the cusp at $\infty$.

Alternatively, we may replace the $q$-series $g$ by the $q$-series $g' = \Delta^j g$ for a positive integer $j$ such that $\Delta^j g$ is a power series. Here $\Delta$ is the normalized cusp form of weight 12. Then one can literally apply [Borcherds 1999, Theorem 3.1] to $g'$ to deduce modularity in $M_{k+12j}(\rho)$ of this power series. Dividing out the power of $\Delta$ again, we obtain the result. □

Proof of Theorem 4.2. According to Proposition 4.3 with $\rho = \omega_L^\vee$, it suffices to show that

$$\sum_{d \in \mathbb{Q}} (c(-d), [Z(d)]_m) = 0 \in J^\text{add}_m(X), \quad (4.12)$$

for every

$$f(\tau) = \sum_{d \in \mathbb{Q}} c(d) \cdot q^d \in M^1_{1/2}(\omega_L). \quad (4.13)$$

Since the space $M^1_{1/2}(\omega_L)$ has a basis of modular forms with integral coefficients [McGraw 2003], it suffices to check that for every $f$ with integral coefficients the relation (4.12) holds. For $\mu \in L'$ we put $c(d, \mu) = c(d)(\mu)$.

Let $\Psi(z, f)$ be the Borcherds lift of $f$ as in [Borcherds 1998, Theorem 13.3]. This is a meromorphic modular form on $\mathcal{D}^+$ for the group $\Gamma$ of weight $c(0, 0)$ with some multiplier system of finite order. Its divisor on $X$ is given by

$$\text{div}(\Psi(z, f)) = \sum_{d > 0} (c(-d), Z(d)) + B(f),$$

where $B(f)$ is a divisor of degree $\frac{1}{12} c(0, 0)$ supported at the cusps of $X$. Let $I \in \text{Iso}(L)$. To determine the behavior of $\Psi(z, f)$ near the cusp $c_I$, we identify $\mathcal{D}^+$ with the upper complex half-plane $\mathbb{H}$ using (3.4). Then $\Psi(w, f)$ has the infinite product expansion

$$\Psi(w, f) = R_I \cdot q^{\rho_I} \prod_{\lambda \in (L' \cap I^\perp) / I \atop n_I(\lambda) > 0} (1 - e^{2\pi i \lambda, \ell'} q^{n_I(\lambda)})^{c(-Q(\lambda), \lambda)}, \quad (4.14)$$

which converges near the cusp $c_I$, that is, for $w \in \mathbb{H}$ with sufficiently large imaginary part. Here the product runs over vectors $\lambda$ of negative norm which are associated to $I$, and $\rho_I \in \mathbb{Q}$ is the Weyl vector at the cusp $c_I$ corresponding to $f$. Moreover, the quantity $R_I$ is some constant in $k^\times$ of modulus 1 times

$$\prod_{a \in \mathbb{Z} / N_I \mathbb{Z} \atop a \neq 0} (1 - e^{2\pi ia/N_I})^{c(0, a\ell/N_I)/2}.$$
Hence, the (finite) product
\[ \Psi(w, f) \times R_I^{-1} \prod_{\lambda \in (L' \cap I^\perp)/I \atop m_I > n_I(\lambda) > 0} h_\lambda^{c(-Q(\lambda), \lambda)} \]
is a meromorphic modular form of weight \( c(0, 0) \) satisfying the condition (2-6) at \( c_I \). There exists a degree-zero divisor \( D \) supported on \( S \) such that the finite product
\[ \Psi(w, f) \times u^{-1}_{D, (R_I)} \times \prod_{I \in \text{Iso}(L)/\Gamma} \prod_{\lambda \in (L' \cap I^\perp)/I \atop m_I > n_I(\lambda) > 0} h_\lambda^{c(-Q(\lambda), \lambda)} \]
is a meromorphic modular form of weight \( c(0, 0) \) satisfying the condition (2-6) at all cusps and having order 0 at all cusps different from \( c_{I_0} \). Here \( u_{D, r} \in H_{G_m, m} \) denotes the function defined in (4-1).

By the choice of the base point \( s_0 = c_{I_0} \) in (2-7), the class of the line bundle \( \mathcal{M}_{c(0,0)} \) in \( J_m(X) \) is given by
\[
[\mathcal{M}_{c(0,0)}]_m = [\text{div}(\Psi(f)) - \deg(\mathcal{M}_{c(0,0)})(c_{I_0})]_m - [\text{div}(u_{D, (R_I)})]_m
- \sum_{I \in \text{Iso}(L)/\Gamma} \sum_{\lambda \in (L' \cap I^\perp)/I \atop m_I > n_I(\lambda) > 0} c(-Q(\lambda), \lambda) \cdot [\text{div}(h_\lambda)]_m. \tag{4-15}
\]
Using Lemma 4.1, we see that
\[
\sum_{I \in \text{Iso}(L)/\Gamma} \sum_{\lambda \in (L' \cap I^\perp)/I \atop m_I > n_I(\lambda) > 0} c(-Q(\lambda), \lambda) \cdot [\text{div}(h_\lambda)]_m = \sum_{d < 0} (c(-d), [Z(d)]_m).
\]
Inserting this into (4-15), we obtain the relation
\[-c(0, 0)[\mathcal{M}_{-1}]_m = \sum_{d > 0} (c(-d), [Z(d)]_m) + \sum_{d < 0} (c(-d), [Z(d)]_m) - [\text{div}(u_{D, (R_I)})]_m \]
in \( J_m(X) \). This implies (4-12) in \( J_m^\text{add}(X) \), concluding the proof. \( \square \)

**Remark 4.4.** To be able to describe the generating series in \( J_m(X) \) instead of in the quotient \( J_m^\text{add}(X) \), we would have to know the normalizing factors \( R_I \) in (4-14) more precisely. It would be very interesting to understand these better. Are they roots of unity?

By the Manin–Drinfeld theorem, the natural homomorphism \( J_m(X) \to J(X) \) induces a linear map
\[
J_m^\text{add}(X) \to J(X).
\]
The classes \([Z(d)]_m\) with \( d \leq 0 \) are in the kernel. Applying this map coefficientwise to the generating series \( A_m \) in Theorem 4.2, we obtain the Gross–Kohnen–Zagier
Corollary 4.5 (Gross–Kohnen–Zagier). The generating series
\[ A_0(\tau) = \sum_{d > 0} [Z(d)]_0 \cdot q^d \]
of the classes of the Heegner divisors in the Jacobian \( J(X)\mathbb{C} \) is the \( q \)-expansion of a cusp form in \( S_{3/2}(\omega^\vee_L) \otimes J(X)\mathbb{C} \).

5. Traces of singular moduli

Here we show that every harmonic Maass form of weight zero with vanishing constant terms defines a linear functional of the generalized Jacobian \( J_{m}^{\text{add}}(X)\mathbb{C} \). Applying it to the generating series \( A_m \), one obtains modularity results for traces of CM values of harmonic Maass forms and weakly holomorphic modular forms as in [Zagier 2002; Bruinier and Funke 2006].

Let \( H^+_k(\Gamma) \) be the space of harmonic Maass forms of weight \( k \) for \( \Gamma \) as in [Bruinier and Funke 2004, Section 3]. Recall that there is a surjective differential operator \( \xi_k : H^+_k(\Gamma) \to S_{2-k}(\Gamma) \) to cusp forms of “dual” weight \( 2 - k \).

For the rest of this section we fix a nonzero \( F \in H^+_0(\Gamma) \). We denote the holomorphic part of the Fourier expansion of \( F \) at the cusp \( c_I \) corresponding to \( I \in \text{Iso}(L) \) by
\[ F^+_I = \sum_{j \in \mathbb{Z}} c^+_F, I (j) \cdot q^j. \quad (5-1) \]

We define the order of \( F \) at the cusp \( c_I \) by
\[ \text{ord}_{c_I}(F) = \min\{ j \in \mathbb{Z} : c^+_F, I (j) \neq 0 \}. \]

Proposition 5.1. Assume that for all \( I \in \text{Iso}(L) \) we have \( \text{ord}_{c_I}(F) > -m_I \) and \( c^+_F, I (0) = 0 \).

(i) There is a linear map,
\[ \text{tr}_F : J_m(X) \to \mathbb{C}, \]
defined by
\[ [D]_m \mapsto \text{tr}_F(D) := \sum_{a \in \text{supp}(D) \setminus S} n_a \cdot F(a), \]
for divisors \( D = \sum_a n_a \cdot (a) \) in \( \text{Div}^0(X) \).

(ii) The map \( \text{tr}_F \) vanishes on \( H_{G_m, m} \) and factors through \( J_m^{\text{add}}(X) \).

Proof. (i) We have to show that \( \text{tr}_F(D) = 0 \), for every divisor \( D = \text{div}(g) \in P_m(X) \) given by a rational function \( g \in k(X)^\times \) satisfying
\[ q_I^{-\text{ord}_{c_I}(g)} g = 1 + O(q_I^{m_I}) \]
at every cusp \( c_I \). The expansion of the logarithmic derivative of \( g \) with respect to the local parameter \( q_I \) at \( c_I \) is of the form

\[
\frac{dg}{g} = \text{ord}_{c_I}(g)q_I^{-1} + O(q_I^{m_I-1}).
\]

If \( F \) is weakly holomorphic, then \( \eta := F(dg/g) \) is a meromorphic 1-form on \( X \). Hence, by the residue theorem, the sum of the residues of \( \eta \) vanishes, and we have

\[
\sum_{a \in \partial X \setminus S} \text{res}_a(\eta) = -\sum_{a \in S} \text{res}_a(\eta).
\]

The left-hand side of this equality is given by \( \text{tr}_F(D) \), while the right-hand side satisfies

\[
\sum_{a \in S} \text{res}_a(\eta) = \sum_{I \in \text{Iso}(L)/\Gamma} \text{res}_{q_I=0} \left( \left( \text{ord}_{c_I}(g)q_I^{-1} + O(q_I^{m_I-1}) \right) \sum_{j > -m_I} c_{F,I}^+(j) \cdot q_I^j \right) = 0.
\]

Here we have also used the fact that \( c_{F,I}^+(0) = 0 \) for all \( I \).

To prove the assertion for general \( F \in H^+_0(\Gamma) \), we let \( X_{\varepsilon} \) be the manifold with boundary obtained from \( X \) by cutting out small oriented discs of radius \( \varepsilon \) around the points in \( \text{supp}(\text{div}(g)) \cup S \). Then for the 1-form \( \eta := F(dg/g) \) it is still true that

\[
\lim_{\varepsilon \to 0} \int_{\partial X_{\varepsilon}} \eta = 0.
\]

Indeed, we have

\[
\int_{\partial X_{\varepsilon}} \eta = \int_{\partial X_{\varepsilon}} F \cdot \partial \log |g|^2 = \int_{X_{\varepsilon}} d(F \cdot \partial \log |g|^2).
\]

Since \( \log |g|^2 \) and \( F \) are harmonic functions on \( X_{\varepsilon} \), we find that

\[
\int_{\partial X_{\varepsilon}} \eta = \int_{X_{\varepsilon}} (\bar{\partial} F) \wedge (\partial \log |g|^2) = -\int_{X_{\varepsilon}} \partial ((\bar{\partial} F) \log |g|^2) = -\int_{\partial X_{\varepsilon}} (\bar{\partial} F) \log |g|^2.
\]

In the latter integral, the differential \( \bar{\partial} F = \bar{\xi}_0(F)d\bar{z} \) is antiholomorphic (hence smooth) on all of \( X \). Since \( \log |g|^2 \) has only logarithmic singularities, the integral vanishes in the limit \( \varepsilon \to 0 \).

On the other hand, a local computation shows that

\[
\lim_{\varepsilon \to 0} \int_{\partial X_{\varepsilon}} \eta = \text{tr}_F(D) + \sum_{I \in \text{Iso}(L)/\Gamma} \text{res}_{c_I} \left( F^+_I \cdot \frac{dg}{g} \right).
\]

The vanishing of the second summand on the right-hand side follows as before, proving that \( \text{tr}_F(D) = 0 \) again.
(ii) Let $u_{D,r}$ be as in (4-1). The same argument shows that $\text{tr}_F(\text{div}(u_{D,r}))$ vanishes and $\text{tr}_F$ factors through $J^\text{add}_m(X)$. □

**Theorem 5.2.** Assume that $\text{ord}_{c_I}(F) > -m_I$ and $c_{F,I}^+(0) = 0$ for all $I \in \text{Iso}(L)$. Then $\text{tr}_F(A_m)$ is a weakly holomorphic modular form in $M^!_{3/2}(\omega_L^\vee)$, and

$$\text{tr}_F(A_m)(\varphi) = \sum_{d < 0} \text{tr}_F([Z(d, \varphi)]_m) \cdot q^d + \text{tr}_F([M_{-1}]_m)\varphi(0) + \sum_{d > 0} F(Y(d, \varphi)) \cdot q^d.$$ 

Moreover, for $d < 0$ the quantity $\text{tr}_F([Z(d, \varphi)]_m)$ is given by the finite sum

$$\text{tr}_F([Z(d, \varphi)]_m) = -\frac{1}{2} \sum_{I \in \text{Iso}(L)/\Gamma} \sum_{\lambda \in L'_{d,I}/\Gamma_I} (\varphi(\lambda) + \varphi(-\lambda)) \cdot \sum_{j \geq 1} e^{2\pi i (\lambda, I'j)} c_{F,I}^+(n_I(\lambda)j).$$

**Proof.** The modularity of $\text{tr}_F(A_m)$ is a direct consequence of Theorem 4.2 and Proposition 5.1.

We now compute the $q$-expansion. For $d > 0$ and $\varphi \in S_L$ we have by definition of the map $\text{tr}_F$ that

$$\text{tr}_F([Z(d, \varphi)]_m) = F(Y(d, \varphi)).$$

If $d < 0$ and $d \in -2 \text{disc}(L)(\mathbb{Q}^\times)^2$, we obtain by the definition of the class $[Z(d)]_m$ that

$$\text{tr}_F([Z(d, \varphi)]_m) = \sum_{\lambda \in L'_{d}/\Gamma} \frac{1}{2n_I(\lambda)} (\varphi(\lambda) + \varphi(-\lambda)) \cdot \text{tr}_F(\text{div}(h_{\lambda}^{-1}))$$

$$= -\sum_{I \in \text{Iso}(L)/\Gamma} \sum_{\lambda \in L'_{d,I}/\Gamma_I} \frac{1}{2n_I(\lambda)} (\varphi(\lambda) + \varphi(-\lambda)) \cdot F(\text{div}(h_{\lambda})).$$

Arguing as in the proof of Proposition 5.1, in particular (5-2), we find for $\lambda \in L'_{d,I}$ that

$$F(\text{div}(h_{\lambda})) = -\sum_{I \in \text{Iso}(L)/\Gamma} \text{res}_c\left(F^+_I \cdot \frac{dh_{\lambda}}{h_{\lambda}}\right)$$

$$= \text{res}_c\left(F^+_I \cdot \frac{n_I(\lambda) \cdot e^{2\pi i (\lambda, I'j)} q_I^{n_I(\lambda)-1} + O(q_I^{m_I-1})}{1 - e^{2\pi i (\lambda, I'j)} q_I^{n_I(\lambda)}}\right)$$

$$= n_I(\lambda) \sum_{j \geq 1} e^{2\pi i (\lambda, I'j)} c_{F,I}^+(n_I(\lambda)j).\quad (5-3)$$

Inserting this into the previous equation, we obtain the assertion. □

**Remark 5.3.** The constant term $\text{tr}_F([Z(0, \varphi)]_m)$ can also be computed explicitly, see Proposition 5.4 for an example.
An example. Consider the modular curve $X_0(M)$ for a squarefree $M \in \mathbb{Z}_{>0}$. Let $L$ be the lattice

$$L = \left\{ \left( \frac{b}{c} \ a/M \ -b \right) : a, b, c \in \mathbb{Z} \right\}, \quad (5-4)$$

with the quadratic form $Q(X) = M \det(X)$. Then $L'/L \cong \mathbb{Z}/2M\mathbb{Z}$ and $\text{SO}^+(L)$ is isomorphic to the extension $\Gamma_0^*(M)$ of $\Gamma_0(M)$ by the Atkin–Lehner involutions. The discriminant kernel subgroup $\Gamma$ is isomorphic to $\Gamma_0^*(M)$, and the modular curve $X_\Gamma$ is isomorphic to $X_0(M)$ with the cusp associated to the isotropic line $I_0 = \mathbb{Z}\left( \begin{smallmatrix} 0 & 0 \\ 1 & 0 \end{smallmatrix} \right)$ corresponding to $\infty$; see, e.g., [Bruinier and Ono 2010, Section 2.4].

The group $\Gamma_0^*(M)$ acts transitively on $\text{Iso}(L)$, and the orbits are represented by the lines $I_D = W_D.I_0$ for the positive divisors $D | M$. Here $W_D \in \text{PGL}_2^+(\mathbb{Q})$ denotes the Atkin–Lehner involution with index $D$. In particular, the set $S$ of cusps $X_0(M)$ is in bijection with the set of positive divisors of $M$. If $I \in \text{Iso}(L)$, we write $D_I$ for the unique positive divisor of $M$ such that $I$ is equivalent to $W_{D_I}.I_0$ under $\Gamma$. Let $F \in H^+_{00}(\Gamma)$ be a harmonic Maass form. The expansion of $F$ at the cusp $I_D$ as in (5-1) is given by the Fourier expansion of $F \mid W_{D_I}$.

Proposition 5.4. Assume that for all $I \in \text{Iso}(L)$ we have $\text{ord}_{c_I}(F) > -m_I$ and $c_{F, I}(0) = 0$. The constant term of the generating series $\text{tr}_F(A_m)$ is given by

$$\text{tr}_F([\mathcal{M}_{-1}]_m) = 2 \sum_{D | M} \sum_{j \geq 1} c_{F, I_D}^+(-j) \cdot D \cdot \sigma_1(j/D).$$

Remark 5.5. As shown in [Bruinier and Funke 2006, Remark 4.9], the right-hand side above is also equal to $-\frac{1}{4\pi} \int_0^{\text{reg}} F d\mu$. The proposition gives a geometric interpretation of this regularized integral.

Proof of Proposition 5.4. We use the notation of the proof of Theorem 5.2. By linearity it suffices to compute the class of the line bundle $\mathcal{M}_{12}$. Since $X_{\Gamma} \cong X_0(M)$, a section of this line bundle is the usual discriminant function given by

$$\Delta = q \prod_{n \geq 1} (1 - q^n)^{24}.$$

To compute the class of $\mathcal{M}_{12}$ in the generalized Jacobian, we have to modify this section by multiplying with rational functions such that the local conditions (2-6) at the cusps are satisfied. It is easily checked that

$$\Delta \mid W_D = D^{-6} \Delta(D\tau) = D^{-6} q^D \prod_{n \geq 1} (1 - q^{Dn})^{24}.$$

This implies that the section

$$s = \Delta \cdot \prod_{I \in \text{Iso}(L)/\Gamma} \prod_{\lambda \in (L'/L)^\perp/I} h_{D_I\lambda}^{-24}$$

for

$$s \mid h_{D_I\lambda}^{-24}.$$
has the expansion
\[ s = D^{-6} q_{I_D}^D \cdot (1 + O(q_{I_D}^{Dm_I})) \]
at the cusp \( I_D \). For the \(|S|\)-tuple \( r = (D^6)_{D|M} \), and the function \( u_{0,r} \in \mathbb{Q}(X_0(M))^\times \), the section \( s \cdot u_{0,r} \) of \( M_{12} \) satisfies the local conditions (2-6) at all cusps. Therefore, in view of (2-7) and Proposition 5.1 (ii), we have
\[
\text{tr}_F([M_{12}]_m) = \text{tr}_F([\text{div}(s \cdot u_{0,r}) - \deg(M_{12}) \cdot (c_{I_0})]_m)
= -24 \sum_{I \in \text{Iso}(L)/\Gamma} \sum_{\lambda \in (L' \cap I^\perp)/I} F(\text{div}(h_{D,I\lambda})).
\]
Using formula (5-3), we get
\[
\text{tr}_F([M_{12}]_m) = -24 \sum_{D|M} \sum_{j \geq 1} c_{F,I_D}^+(-j) \cdot D \cdot \sigma_1(j/D).
\]
This concludes the proof of the proposition. \( \square \)

We now explain how to obtain a scalar-valued generating series from \( \text{tr}_F(A_m) \). By means of the canonical pairing \( (S_L, S_L^\vee) \rightarrow \mathbb{C} \), we define a map
\[
S_L^\vee \rightarrow \mathbb{C}, \quad u \mapsto (\chi_1, u),
\]
given by the pairing with the constant function \( \chi_1 \) with value 1. It induces a map from \( S_L^\vee \)-valued to scalar-valued modular forms,
\[
M_{3/2}(\omega_L^\vee) \rightarrow M_{3/2}(\Gamma_0(4M)), \quad f(\tau) \mapsto f^{\text{scal}}(\tau) := f(\chi_1)(4M \tau),
\]
see [Eichler and Zagier 1985, §5]. The image lies in the Kohnen plus-space. Applying this map to the generating series \( A_m \) of Theorem 4.2, we obtain a scalar valued generating series which has level \( 4M \). In particular, this implies Theorem 1.1 of the introduction. If we apply this map to Theorem 5.2 and use Proposition 5.4, we obtain:

**Theorem 5.6.** Let \( L \) be as in (5-4). Assume that for all \( I \in \text{Iso}(L) \) we have \( \text{ord}_{c_I}(F) > -m_I \) and \( c_{F,I}^+(0) = 0 \). Then \( \text{tr}_F(A_m^{\text{scal}}) \in M_{3/2}^!(\Gamma_0(4M)) \), and
\[
\text{tr}_F(A_m^{\text{scal}}) = - \sum_{D|M} \sum_{b \geq 1} \sum_{n \geq 1} c_{F,I_D}^+(-bn) \cdot b \cdot q^{-b^2} \\
+ 2 \sum_{D|M} \sum_{n \geq 1} c_{F,I_D}^+(-n) \cdot D \cdot \sigma_1(n/D) + \sum_{d \in \mathbb{Z}_{>0}} F(Y(d/4M, \chi_1)) \cdot q^d.
\]
When $M = p$ is a prime and $F$ is invariant under the Fricke involution, we obtain Theorem 1.4 of the introduction.

Now let $M = 1$ and let $j = E_4^3/\Delta$ be the classical $j$-function. Write $J = j - 744 = q^{-1} + 196884q + \cdots$ for the normalized Hauptmodul for $\text{PSL}_2(\mathbb{Z})$ with vanishing constant term. Applying Theorem 5.6 with $F = J$, we recover Zagier’s original result [2002]:

**Corollary 5.7.** The generating series

$$-q^{-1} + 2 + \sum_{d \in \mathbb{Z}_{d > 0}} J(Y(d/4, \chi_1)) \cdot q^d$$

of the traces of singular moduli is a weakly holomorphic modular form for $\Gamma_0(4)$ of weight $3/2$ in the plus-space.

6. Generalizations

In the section we describe some variants of our main results and indicate possible generalizations.

**Modularity in the generalized class group.** In the definition of the Heegner divisors $Z(d)$ we have projected to degree-0 divisors by subtracting a suitable multiple of $(cI_0)$. We now briefly describe what happens if we do not apply this projection and consider the divisors $Y(d, \varphi)$ defined in (3-7) for $d > 0$. Then the corresponding generating series is a nonholomorphic modular form, where the nonholomorphic part is coming from a generalization of Zagier’s weight-$\frac{3}{2}$ Eisenstein series.

We let $\text{Cl}_m(X)$ be the generalized class group of $X$ with respect to the modulus $m$, which we define as the quotient of the group of divisors on $X$ defined over $k$ modulo $P_m(X)$. Moreover, in analogy with (4-2) we put

$$\text{Cl}^{\text{add}}_m(X) = \text{Cl}_m(X)/H_{G_m,m}. \quad (6-1)$$

If $d > 0$, we write $[Y(d, \varphi)]_m$ for the class of the divisor $Y(d, \varphi)$ in $\text{Cl}_m(X)$. For $d = 0$ we put $[Y(0, \varphi)]_m = \varphi(0)[M_{-1}]_m$, where the class in $\text{Cl}_m(X)$ of a line bundle $\mathcal{L}$ is defined as in (2-7) but without the summand $\text{deg} (\mathcal{L}) \cdot (s_0)$. Finally, for $d < 0$ we let $[Y(d, \varphi)]_m = [Z(d, \varphi)]_m$.

Recall from [Funke 2002, Theorem 3.5] that there is a (nonholomorphic) weight-$\frac{3}{2}$ Eisenstein series $E_{3/2,L}(\tau)$ whose coefficients with nonnegative index are given by the degrees of the $Y(d, \varphi)$ (see also [Kudla 2003]). It is a harmonic Maass form of weight $\frac{3}{2}$ for the group $\text{Mp}_2(\mathbb{Z})$ with representation $\omega_L^\vee$ and generalizes Zagier’s nonholomorphic Eisenstein series [1975]. Its Fourier expansion decomposes as

$$E_{3/2,L}(\tau) = E_{3/2,L}^+, (\tau) + E_{3/2,L}^-(\tau),$$
where the holomorphic part is the generating series of the degrees of Heegner divisors,
\[ E_{3/2,L}^+(\tau) = \sum_{d \geq 0} \deg(Y(d)) \cdot q^d, \]
and the nonholomorphic part \( E_{3/2,L}^- \) is a period integral of a linear combination of unary theta series. We obtain the following variant of Theorem 4.2.

**Theorem 6.1.** The generating series
\[ \tilde{A}_m(\tau) = \sum_{d \in \frac{1}{2} \mathbb{Z}} [Y(d)]_m \cdot q^d + E_{3/2,L}^-(c_{l_0}) \]
is a nonholomorphic modular form of weight \( \frac{3}{2} \) for \( \text{Mp}_2(\mathbb{Z}) \) with representation \( \omega_L^{\vee} \) with values in \( \text{Cl}^{\text{add}}_m(X) \). Moreover, we have
\[ A_m = \tilde{A}_m - E_{3/2,L}^- \cdot (c_{l_0}). \]

**Twists by genus characters.** Let \( L \) be the lattice of page 1294 for a squarefree \( M \in \mathbb{Z}_{>0} \), and recall that \( \Gamma \cong \Gamma_0(M) \). For a discriminant \( \Delta \neq 1 \) and \( r \in \mathbb{Z} \) such that \( \Delta \equiv r^2 \mod 4M \), we can define a generalized genus character \( \chi_\Delta \) on \( L' \) as in [Gross et al. 1987, Section I.2] and [Bruinier and Ono 2010, Section 4] let
\[ \chi_\Delta(\lambda) = \begin{cases} \left( \frac{\Delta}{n} \right) & \text{if } \Delta | b^2 - 4Mc \text{ and } (b^2 - 4Mc)/\Delta \text{ is a square modulo } 4M \text{ and } \gcd(a, b, c, \Delta) = 1, \\ 0 & \text{otherwise}, \end{cases} \]
with \( \lambda = \left( \frac{b/2M}{c} \right) \) \( \in L' \) and \( n \in \mathbb{Z} \) any integer prime to \( \Delta \) represented by one of the quadratic forms \( [M_1 a, b, M_2 c] \) with \( M_1 M_2 = M \). Note that \( \chi_\Delta \) is invariant under \( \text{SO}^+(L) \).

If \( \lambda \in L' \) with \( Q(\lambda) \in -4M(\mathbb{Q}^\times)^2 \), let \( I \) be the isotropic line associated with \( \lambda \), and let \( h_{\Delta,\lambda} \in \mathbb{Q}((\sqrt{\Delta})(X)^\times \) be a rational function with the following expansions:
\[ h_{\Delta,\lambda} = \prod_{b \in \mathbb{Z}/\Delta} (1 - e^{-2\pi i b/\Delta} q_I^{n_{I}(\lambda)})^{\left( \frac{\Delta}{2} \right)} + O(q_I^{m_{I}}) \]
at the cusp \( c_I \),
\[ h_{\Delta,\lambda} = 1 + O(q_{J}^{m_{J}}) \]
at all other cusps \( c_J \).

Suppose that \( (\Delta, 2M) = 1 \), or equivalently \( (r, 2M) = 1 \). For each \( d \in \frac{1}{4M} \mathbb{Z} \) and \( \varphi \in S_L \), we can define the divisor \( Z_{\Delta,r}(d, \varphi) \in \text{Div}^0(X) \) by
\[ Z_{\Delta,r}(d, \varphi) := \begin{cases} \sum_{\lambda \in L_{d/\Delta}/\Gamma} \frac{\chi_\Delta(\lambda) \varphi(r^{-1} \lambda)}{2|\Gamma|} \cdot (z_\lambda) & \text{if } d > 0, \\ \sum_{\lambda \in L'_{d/\Delta}/\Gamma} \frac{\varphi(r \lambda) + \text{sgn}(\Delta) \varphi(-r \lambda)}{2n_{I}(\lambda)} \text{div}(h_{\Delta,\lambda}^{-1}) & \text{if } d \in -\frac{|\Delta|}{4M}(\mathbb{Z}_{>0})^2, \\ 0 & \text{otherwise}. \end{cases} \]
All these divisors are defined over $\mathbb{Q}(\sqrt{\Delta})$ and have coefficients in the field of definition of $\varphi$. We write $[Z_{\Delta,r}(d)]_m \in S^\vee_L \otimes J^\text{add}_m(X)$ for the element that sends $\varphi$ to $[Z_{\Delta,r}(d, \varphi)]_m \in J^\text{add}_m(X)$. Define the representation $\tilde{\omega}_L$ to be $\omega_L$ if $\Delta > 0$ and $\tilde{\omega}_L$ if $\Delta < 0$. Then we have the following abstract modularity result.

**Theorem 6.2.** The generating series

$$A_{\Delta,r,m}(\tau) := \sum_{d \in \mathbb{Z}^+} [Z_{\Delta,r}(d)]_m \cdot q^d \in S^\vee_L((q)) \otimes J^\text{add}_m(X)_C$$

is the $q$-expansion of a weakly holomorphic modular form in $M^!_{3/2}(\tilde{\omega}_L) \otimes J^\text{add}_m(X)_C$.

This result comes out of calculating the effect of the intertwining operator in [Alfes and Ehlen 2013, Section 3] applied to the generating series $A_m(\tau)$ associated to the scaled lattice $(\Delta L, Q(\cdot)/|\Delta|)$. The conditions that $M$ is squarefree and $(\Delta, 2M) = 1$ are imposed to simplify the definition of $Z_{\Delta,r}(d, \varphi)$ and can be removed with a more complicated definition of the classes. Note that it is necessary for $\text{sgn}(\Delta)$, which determines the parity of $\tilde{\omega}_L$, to appear in the definition of $Z_{\Delta,r}(d, \varphi)$. Alternatively, one could use the twisted Borcherds products in [Bruinier and Ono 2010, Theorem 6.1] to give a proof of Theorem 6.2 along the same line as that of Theorem 4.2 above. By applying the functionals of Proposition 5.1 to the twisted generating series of Theorem 6.2, the main result of [Alfes and Ehlen 2013] on twisted traces of harmonic Maass forms can be recovered.

**Other orthogonal Shimura varieties.** The Gross–Kohnen–Zagier theorem has been generalized to higher dimensional orthogonal Shimura varieties in [Borcherds 1999]. Hence it is natural to ask whether our main results can also be generalized in the same direction. Let $L$ be an even lattice of signature $(n, 2)$, and let $\Gamma$ be the discriminant kernel subgroup of $\text{SO}^+(L)$. Denote by $X_\Gamma$ a (suitable) toroidal compactification of the connected Shimura variety $Y_\Gamma$ associated to $\Gamma$. It would be interesting to define a generalized divisor class group as the group of divisors on $X_\Gamma$ modulo divisors of rational functions that satisfy certain growth conditions along the boundary of $X_\Gamma$. Is it possible to prove a modularity result analogous to Theorem 4.2 for the classes of special divisors? In this context, the product expansions obtained in [Kudla 2014] with respect to one-dimensional Baily–Borel boundary components may be helpful.

To illustrate this question, let us consider the easiest case for $n = 2$ where the lattice $L$ is the even unimodular lattice of signature $(2, 2)$. Then the variety $X_\Gamma$ can be identified with the product $X(1) \times X(1)$ of two copies of the compact modular curve of level 1. Special divisors on $X_\Gamma$ of positive index $d$ in the sense of [Kudla 1997] are given by the Hecke correspondences $Z(d)$. Let $q = (q_1, q_2)$ be the usual local coordinates near the boundary point $s = (\infty, \infty) \in X_\Gamma$. Let $m$ be a positive
integer, and put $m = m \cdot (s)$. If $k = (k_1, k_2) \in \mathbb{Z}^2$ we briefly write $q^k = q_1^{k_1} q_2^{k_2}$, and for a meromorphic function $f$ in a neighborhood of $(\infty, \infty)$ we write $f = O(q^m)$ if in the Taylor expansion of $f$ at $(\infty, \infty)$ only terms of total degree at least $m$ occur.

Let $\text{Div}_m(X_\Gamma)$ be the free abelian group generated by pairs $(D, g_D)$, where $D$ is a prime Weil divisor on $X_\Gamma$ and $g_D$ is a local equation for $D$ in a small neighborhood of $s$. The local equations give rise to local equations $g_D$ near $s$ for arbitrary Weil divisors $D$. Let $P_m(X_\Gamma)$ be the subgroup of pairs $(D, g_D) \in \text{Div}_m(X_\Gamma)$ for which $D = \text{div}(f)$ is the divisor of a meromorphic function $f$ satisfying

$$f \cdot g_D^{-1} = 1 + O(q^m)$$

near $s$. We define a generalized class group as the quotient

$$\text{Cl}_m(X_\Gamma) = \text{Div}_m(X_\Gamma) / P_m(X_\Gamma).$$

It would be interesting to define suitable classes of special divisors in $\text{Cl}_m(X_\Gamma)$ of arbitrary integral index $d$ and to prove a modularity result for the generating series of these classes.

Acknowledgments

We thank J. Funke, B. Gross, and S. Kudla for useful conversations on the content of this paper. Moreover, we thank the referee for his/her valuable comments.

References


Communicated by Richard Borcherds

Received 2015-08-27 Revised 2016-04-20 Accepted 2016-05-19

bruinier@mathematik.tu-darmstadt.de

Fachbereich Mathematik, Technische Universität Darmstadt, Schloßgartenstrasse 7, D-64289 Darmstadt, Germany

li@mathematik.tu-darmstadt.de

Fachbereich Mathematik, Technische Universität Darmstadt, Schloßgartenstrasse 7, D-64289 Darmstadt, Germany
On 2-dimensional 2-adic Galois representations of local and global fields

Vytautas Paškūnas

We describe the generic blocks in the category of smooth locally admissible mod-2 representations of $GL_2(\mathbb{Q}_2)$. As an application we obtain new cases of the Breuil–Mézard and Fontaine–Mazur conjectures for 2-dimensional 2-adic Galois representations.

1. Introduction

Let $p$ be a prime and let $L$ be a finite extension of $\mathbb{Q}_p$ with the ring of integers $\mathcal{O}$ and uniformizer $\varpi$. We prove the following modularity lifting theorem.

**MSC2010:** 11F80.

**Keywords:** $p$-adic Langlands, Fontaine–Mazur, modularity lifting.
Theorem 1.1. Assume that \( p = 2 \). Let \( F \) be a totally real field where 2 is totally split, let \( S \) be a finite set of places of \( F \) containing all the places above 2 and all the infinite places and let

\[
\rho : G_{F,S} \to \text{GL}_2(\mathcal{O})
\]

be a continuous representation of the Galois group of the maximal extension of \( F \) unramified outside \( S \). Suppose:

(i) \( \overline{\rho} : G_{F,S} \to \text{GL}_2(\mathcal{O}) \to \text{GL}_2(k) \) is modular with nonsolvable image.

(ii) If \( v \mid 2 \) then \( \rho|_{G_F} \) is potentially semistable with distinct Hodge–Tate weights.

(iii) \( \text{det} \rho \) is totally odd.

(iv) If \( v \mid 2 \) then \( \overline{\rho}|_{G_F} \not\sim (\chi \otimes \chi) \) for any character \( \chi : G_{F_v} \to k^\times \).

Then \( \rho \) is modular.

Kisin [2009a] and Emerton [2011] have proved an analogous theorem for \( p > 2 \). Our proof follows the strategy of Kisin. We patch automorphic forms on definite quaternion algebras and deduce the theorem from a weak form of the Breuil–Mézard conjecture, which we prove for all \( p \) under some technical assumptions on the residual representation of \( G_{\mathbb{Q}_p} \) (see Theorems 2.34 and 2.37) which force us to assume (iv) in the theorem.

The Breuil–Mézard conjecture is proved by employing a formalism developed in [Paškūnas 2015b], where an analogous result is proved under the assumption that \( p \geq 5 \) and the residual representation has scalar endomorphisms. We can prove the result for primes 2 and 3 by better understanding the smooth representation theory of \( G := \text{GL}_2(\mathbb{Q}_p) \) in characteristic \( p \): in the local part of the paper we extend the results of [Paškūnas 2013] to the generic blocks, when \( p \) is 2 and 3, which we will now describe.

Let \( \text{Mod}^\text{sm}_G(\mathcal{O}) \) be the category of smooth \( G \)-representation on \( \mathcal{O} \)-torsion modules. We fix a continuous character \( \psi : \mathbb{Q}_p^\times \to \mathcal{O}^\times \) and let \( \text{Mod}^{\text{ladm}}_{G,\psi}(\mathcal{O}) \) be the full subcategory of \( \text{Mod}^\text{sm}_G(\mathcal{O}) \), consisting of representations on which the center of \( G \) acts by the character \( \psi \) and which are equal to the union of their admissible subrepresentations. The categories \( \text{Mod}^\text{sm}_G(\mathcal{O}) \) and \( \text{Mod}^{\text{ladm}}_{G,\psi}(\mathcal{O}) \) are abelian; see [Emerton 2010a, Proposition 2.2.18]. A finitely generated smooth admissible representation of \( G \) with a central character is of finite length by Theorem 2.3.8 of [Emerton 2010a]. This makes \( \text{Mod}^{\text{ladm}}_{G,\psi}(\mathcal{O}) \) into a locally finite category. Gabriel [1962] has proved that a locally finite category decomposes into a direct product of indecomposable subcategories as follows.

Let \( \text{Irr}^{\text{ladm}}_G \) be the set of irreducible representations in \( \text{Mod}^{\text{ladm}}_{G,\psi}(\mathcal{O}) \). We define an equivalence relation \( \sim \) on \( \text{Irr}^{\text{ladm}}_G \) by writing \( \pi \sim \tau \) if there exists a sequence \( \pi = \pi_1, \pi_2, \ldots, \pi_n = \tau \) in \( \text{Irr}^{\text{ladm}}_G \) such that for each \( i \) one of the following holds:
(1) $\pi_i \cong \pi_{i+1}$; (2) $\text{Ext}^1_G(\pi_i, \pi_{i+1}) \neq 0$; (3) $\text{Ext}^1_G(\pi_{i+1}, \pi_i) \neq 0$. We have a canonical decomposition

$$\text{Mod}^{\text{ladm}}_{G, \psi}(\mathcal{O}) \cong \coprod_{\mathcal{B} \in \text{Irr}^{\text{ladm}}_G/\sim} \text{Mod}^{\text{ladm}}_{G, \psi}(\mathcal{O})[\mathcal{B}],$$

(1)

where $\text{Mod}^{\text{ladm}}_{G, \psi}(\mathcal{O})[\mathcal{B}]$ is the full subcategory of $\text{Mod}^{\text{ladm}}_{G, \psi}(\mathcal{O})$ consisting of representations with all irreducible subquotients in $\mathcal{B}$. A block is an equivalence class of $\sim$.

For a block $\mathcal{B}$ let $\pi_{\mathcal{B}} = \bigoplus_{\pi \in \mathcal{B}} \pi$, let $\pi_{\mathcal{B}} \hookrightarrow J_{\mathcal{B}}$ be an injective envelope of $\pi_{\mathcal{B}}$ and let $E_{\mathcal{B}} := \text{End}_G(J_{\mathcal{B}})$. Then $J_{\mathcal{B}}$ is an injective generator for $\text{Mod}^{\text{ladm}}_{G, \psi}(\mathcal{O})[\mathcal{B}]$, $E_{\mathcal{B}}$ is a pseudocompact ring and the functor $\kappa \mapsto \text{Hom}_G(\kappa, J_{\mathcal{B}})$ induces an antiequivalence of categories between $\text{Mod}^{\text{ladm}}_{G, \psi}(\mathcal{O})[\mathcal{B}]$ and the category of right pseudocompact $E_{\mathcal{B}}$-modules. The inverse functor is given by $m \mapsto (m \widehat{\otimes} E_{\mathcal{B}}, J_{\mathcal{B}}^\vee)$, where $\vee$ denotes the Pontryagin dual; see [Gabriel 1962, Chapitre IV, §4]. The main result of [Paškūnas 2013] computes the rings $E_{\mathcal{B}}$ for each block $\mathcal{B}$ and describes the Galois representation of $G_{\mathbb{Q}_p}$ obtained by applying the Colmez’s functor to $J_{\mathcal{B}}$ under the assumption $p \geq 5$ or $p \geq 3$, depending on the block $\mathcal{B}$.

If $\pi \in \text{Irr}^{\text{ladm}}_G$ then one may show that, after extending scalars, $\pi$ is isomorphic to a finite direct sum of absolutely irreducible representations of $G$. It has been proved in [Paškūnas 2014] that the blocks containing an absolutely irreducible representation are given by

(i) $\mathcal{B} = \{\pi\}$ with $\pi$ supersingular;

(ii) $\mathcal{B} = \{(\text{Ind}_B^G \chi_1 \otimes \chi_2 \omega^{-1})_{\text{sm}}, (\text{Ind}_B^G \chi_2 \otimes \chi_1 \omega^{-1})_{\text{sm}}\}$ with $\chi_2 \chi_1^{-1} \neq \omega^{\pm 1}, 1$;

(iii) $p > 2$ and $\mathcal{B} = \{(\text{Ind}_B^G \chi \otimes \chi \omega^{-1})_{\text{sm}}\}$;

(iv) $p = 2$ and $\mathcal{B} = \{1, \text{Sp}\} \otimes \chi \circ \det$;

(v) $p \geq 5$ and $\mathcal{B} = \{1, \text{Sp}, (\text{Ind}_B^G \omega \otimes \omega^{-1})_{\text{sm}}\} \otimes \chi \circ \det$;

(vi) $p = 3$ and $\mathcal{B} = \{1, \text{Sp}, \omega \circ \det, \text{Sp} \otimes \omega \circ \det\} \otimes \chi \circ \det$;

where $\chi, \chi_1, \chi_2 : \mathbb{Q}_p^\times \to k^\times$ are smooth characters, $\omega : \mathbb{Q}_p^\times \to k^\times$ is the character $\omega(x) = x|\chi| (\text{mod } \mathfrak{m})$ and we view $\chi_1 \otimes \chi_2$ as a character of the subgroup of upper-triangular matrices $B$ in $G$ which sends $\begin{pmatrix} a & b \\ 0 & d \end{pmatrix}$ to $\chi_1(a) \chi_2(d)$. An absolutely irreducible representation $\pi$ is supersingular if it is not a subquotient of a principal series representation (they have been classified by Breuil [2003a]) and $\text{Sp}$ denotes the Steinberg representation.

To each block above one may attach a semisimple 2-dimensional $k$-representation $\bar{\rho}^{ss}$ of $G_{\mathbb{Q}_p}$; in case (i) $\bar{\rho}^{ss}$ is absolutely irreducible, and such that Colmez’s functor $V$ (see Section 2B1) maps $\pi$ to $\bar{\rho}^{ss}$; in case (ii) $\bar{\rho}^{ss} = \chi_1 \oplus \chi_2$; in cases (iii) and (iv) $\bar{\rho}^{ss} = \chi \oplus \chi$; in cases (v) and (vi) $\bar{\rho}^{ss} = \chi \oplus \chi \omega$, where we consider characters of $G_{\mathbb{Q}_p}$ as characters of $\mathbb{Q}_p^\times$ via local class field theory, normalized so that uniformizers
correspond to geometric Frobenii. We note that the determinant of $\bar{\rho}^{ss}$ is equal to $\psi \varepsilon$ modulo $\omega$, where $\varepsilon$ is the $p$-adic cyclotomic character and $\omega$ is its reduction modulo $\sigma$.

**Theorem 1.2.** If $\mathcal{B} = \{\pi\}$ with $\pi$ supersingular (so that $\bar{\rho}^{ss}$ is irreducible) then $E_\mathcal{B}$ is naturally isomorphic to the quotient of the universal deformation ring of $\bar{\rho}^{ss}$ parametrizing deformations with determinant $\psi \varepsilon$, and $V(J_\mathcal{B})^\vee(\psi \varepsilon)$ is a tautological deformation of $\bar{\rho}^{ss}$ to $E_\mathcal{B}$.

We also obtain an analogous result for blocks in (ii); see Theorem 2.23. Let $R^{ps}$ be the deformation ring parametrizing all the 2-dimensional determinants, in the sense of [Chenevier 2014], lifting $(\text{tr} \bar{\rho}^{ss}, \det \bar{\rho}^{ss})$, and let $R^{ps, \psi}$ be the quotient of $R^{ps}$ parametrizing those which have determinant $\psi \varepsilon$.

**Theorem 1.3.** Assume that the block $\mathcal{B}$ is given by (i) or (ii) above. Then the center of the category $\text{Mod}^{\text{ad}}_G(L)[\mathcal{B}]$ is naturally isomorphic to $R^{ps, \psi}$.

We view this theorem as an analogue of the Bernstein center for this category. Theorems 1.2 and 1.3 are new if $p = 2$ and if $p = 3$ and $\mathcal{B} = \{\pi\}$ with $\pi$ supersingular. Together with the results of [Paškūnas 2013] this covers all the blocks except for those in (iv) and (vi) above.

One also has a decomposition similar to (1) for the category $\text{Ban}^{\text{ad}}_G(L)$ of admissible unitary $L$-Banach space representations of $G$ on which the center of $G$ acts by $\psi$; see Section 2B4. An admissible unitary $L$-Banach space representation $\Pi$ lies in $\text{Ban}^{\text{ad}}_G(L)[\mathcal{B}]$ if and only if all the irreducible subquotients of the reduction modulo $\sigma$ of a unit ball in $\Pi$ modulo $\sigma$ lie in $\mathcal{B}$. An irreducible $\Pi$ is ordinary if it is a subquotient of a unitary parabolic induction of a unitary character. Otherwise it is called nonordinary.

**Corollary 1.4.** Assume that the block $\mathcal{B}$ is given by (i) or (ii) above. Colmez’s Montreal functor $\Pi \mapsto \tilde{V}(\Pi)$ induces a bijection between the isomorphism classes of

- absolutely irreducible nonordinary $\Pi \in \text{Ban}^{\text{ad}}_G(L)[\mathcal{B}]$;
- absolutely irreducible $\tilde{\rho} : G_{\mathbb{Q}_p} \to \text{GL}_2(L)$ such that $\det \tilde{\rho} = \psi \varepsilon$ and the semisimplification of the reduction modulo $\sigma$ of a $G_{\mathbb{Q}_p}$-invariant $O$-lattice in $\tilde{\rho}$ is isomorphic to $\bar{\rho}^{ss}$.

A stronger result, avoiding the assumption on $\mathcal{B}$, is proved in [Colmez et al. 2014]. However, our proof of Corollary 1.4 avoids the hard $p$-adic functional analysis, which is used to construct representations of $\text{GL}_2(\mathbb{Q}_p)$ out of 2-dimensional representations of $G_{\mathbb{Q}_p}$ via the theory of $(\varphi, \Gamma)$-modules by Colmez [2010], which plays the key role in [Colmez et al. 2014].

It might be possible, given the global part of this paper, and the results of [Paškūnas 2015a], where various deformation rings are computed, when $p = 2$,
to prove Theorem 1.1 by repeating the arguments of Kisin [2009a]. We have not checked this. However, our original goal was to prove Theorems 1.2 and 1.3; Theorem 1.1 came out as a bonus at the end.

1A. Outline of the paper. The paper has two largely independent parts: a local one and a global one. We will review each of them individually by carefully explaining which arguments are new.

1A1. Local part. For concreteness, assume that $\mathcal{B} = \{\pi\}$ with $\pi$ supersingular. Let $\bar{\rho} = V(\pi)$, let $R_{\bar{\rho}}$ be the universal deformation ring of $\bar{\rho}$ and let $R_{\bar{\rho}}^\psi$ be the quotient of $R_{\bar{\rho}}$ parametrizing deformations with determinant $\psi$. We follow the strategy outlined in [Paškūnas 2013, §5.8]. We show that $J^{\vee}_{\mathcal{B}}$ is the universal deformation of $\pi^{\vee}$ and $E_{\mathcal{B}}$ is the universal deformation ring by verifying that hypotheses (H0)–(H5), made in Section 3 of [Paškūnas 2013], hold. In Section 3.3 of the same work we developed a criterion to check that the ring $E_{\mathcal{B}}$ is commutative. To apply this criterion, one needs the ring $R_{\bar{\rho}}^\psi$ to be formally smooth and to control the image of some $\text{Ext}^1$-group in some $\text{Ext}^2$-group. The first condition does not hold if $p = 2$ and if $p = 3$ and $\bar{\rho} \cong \bar{\rho} \otimes \omega$. Even if $p = 3$ and $\bar{\rho} \not\cong \bar{\rho} \otimes \omega$, so that the ring is formally smooth, to check the second condition is a computational nightmare. In [Colmez et al. 2014] we found a different characteristic-0 argument to get around this. The key input is the result of [Berger and Breuil 2010] which says that if a locally algebraic principal series representation admits a $G$-invariant norm, then its completion is irreducible, and $\pi$ occurs in the reduction modulo $\sigma$ with multiplicity one. We deduce from [Colmez et al. 2014, Corollary 2.22] that the ring $E_{\mathcal{B}}$ is commutative. The argument of Kisin [2010] shows that $V(J_{\mathcal{B}})^{\vee}(\psi \varepsilon)$ is a deformation of $\bar{\rho}$ to $E_{\mathcal{B}}$ and we have surjections $R_{\bar{\rho}} \twoheadrightarrow E_{\mathcal{B}} \twoheadrightarrow R_{\bar{\rho}}^\psi$.

To prove Theorem 1.2 we have to show that the surjection $\varphi : E_{\mathcal{B}} \twoheadrightarrow R_{\bar{\rho}}^\psi$ is an isomorphism. The proof of this claim is new and is carried out in Section 2B3. Corollary 1.4 is then a formal consequence of this isomorphism. If $p \geq 5$ then $R_{\bar{\rho}}^\psi$ is formally smooth and the claim is proved by a calculation on tangent spaces in [Paškūnas 2013]. This does not hold if $p = 2$ or $p = 3$ and $\bar{\rho} \cong \bar{\rho} \otimes \omega$. We also note that even if we admit the main result of [Colmez et al. 2014] (which we don’t), we would only get that $\varphi$ induces a bijection on maximal spectra of the generic fibers of the rings. From this one could deduce that the map induces an isomorphism between the maximal reduced quotient of $E_{\mathcal{B}}$ and $R_{\bar{\rho}}^\psi$, and it is not at all clear that $E_{\mathcal{B}}$ is reduced. However, by using techniques of [Paškūnas 2015b] we can show that certain quotients $E_{\mathcal{B}}/a$ are reduced and identify them with crystabelline deformation rings of $\bar{\rho}$ via $\varphi$. Again the argument uses the results of [Berger and Breuil 2010] in a crucial way. Further, we show that the intersection of all such ideals in $E_{\mathcal{B}}$ is zero, which allows us to conclude the proof. A similar argument using density appears in [Colmez et al. 2014, §2.4], however we have to work a bit
more here, because we fix a central character; see Section 2A. Theorem 1.2 implies immediately that \( \det \tilde{V}(\Pi) = \psi \varepsilon \) for all \( \Pi \in \text{Ban}_{G,\psi}^{\text{adm}}(L)[2] \). This is proved directly in [Colmez et al. 2014] without any restriction on \( \mathfrak{B} \), and is the most technical part of that paper.

Once we have Theorem 1.2, the Breuil–Mézard conjecture is proved the same way as in [Paškūnas 2015b]; see Section 2C. If \( \mathfrak{B} \) is the block containing two generic principal series representations, so that \( \bar{\rho}^{\text{ss}} = \chi_1 \oplus \chi_2 \), with \( \chi_1 \chi_2^{-1} \neq 1, \omega \pm 1 \), then we prove the Breuil–Mézard conjecture for both nonsplit extensions \( \left( \begin{array}{cc} \chi_1 & * \\ 0 & \chi_2 \end{array} \right) \) and \( \left( \begin{array}{cc} * & \chi_2 \\ \chi_1 & 0 \end{array} \right) \) and deduce the conjecture in the split case in a companion paper [Paškūnas 2015a], following an idea of Hu and Tan [2015]. We formulate and prove the Breuil–Mézard conjecture in the language of cycles, as introduced by Emerton and Gee [2014]. All our arguments are local, except that if the inertial type extends to an irreducible representation of the Weil group \( W_{\mathbb{Q}_p} \) of \( \mathbb{Q}_p \), the description of locally algebraic vectors in the Banach space representations relies on a global input of Emerton [2011, §7.4]. Dospinescu’s results [2015] on locally algebraic vectors in extensions of Banach space representations of \( G \) are also crucial in this case.

1A2. Global part. As already explained, an analogue of Theorem 1.1 has been proved by Kisin if \( p > 2 \). Moreover, if \( p = 2 \) and \( \rho \mid_{\tilde{G}_v} \) is semistable with Hodge–Tate weights \((0, 1)\) for all \( v \mid 2 \), then the theorem has been proved by Khare and Wintenberger [2009b] and Kisin [2009b] in their work on Serre’s conjecture. We use their results as an input in our proof.

The strategy of the proof is the same as in [Kisin 2009a]. By base change arguments, which are the same as in [Khare and Wintenberger 2009b; Kisin 2009b; 2009c] (see Section 3F) we reduce ourselves to a situation where the ramification of \( \rho \) and \( \bar{\rho} \) outside 2 is minimal and \( \bar{\rho} \) comes from an automorphic form on a definite quaternion algebra. We patch automorphic forms on definite quaternion algebras and deduce the theorem from a weak form of the Breuil–Mézard conjecture, which is proved in the local part of the paper. Assumption (iv) in Theorem 1.1 comes from the local part of the paper.

Let us explain some differences with [Kisin 2009a]. If \( p > 2 \) then the patched ring is formally smooth over a completed tensor product of local deformation rings. This implies that the patched ring is reduced, equidimensional and \( \mathcal{O} \)-flat and that its Hilbert–Samuel multiplicity is equal to the product of Hilbert–Samuel multiplicities of the local deformation rings. For \( p = 2 \) we modify the patching argument used in [Kisin 2009a] following [Khare and Wintenberger 2009b]. This gives us two patched rings, and the passage between them and the completed tensor product of local rings is not as straightforward as before. To overcome this we use an idea which appears in errata to [Kisin 2009a] published in [Gee and Kisin 2014]. If \( \rho_f \) is a Galois representation associated to a Hilbert modular form lifting \( \bar{\rho} \) and \( v \) is a place of \( F \) above \( p \), then one knows from [Blasius 2006; Katz and Messing...
On 2-dimensional 2-adic Galois representations of local and global fields 1307

1974; Saito 2009] that the Weil–Deligne representation associated to \( \rho|_{\mathbb{G}_{F_v}} \) is pure. Kisin shows that this implies that the point on the generic fiber of the potentially semistable deformation ring, defined by \( \rho_f|_{\mathbb{G}_{F_v}} \), cannot lie on the intersection of two irreducible components, and hence is regular. Using this we show that the localization of patched rings at the prime ideal defined by \( \rho_f \) is regular, and we are in a position to use the Auslander–Buchsbaum theorem; see Lemma 3.14 and Proposition 3.17. As explained in [Gee and Kisin 2014], this observation enables us to deal with cases when the patched module is not generically free of rank 1 over the patched ring, which was the case in the original paper [Kisin 2009a]. In particular, we don’t add any Hecke operators at places above 2 and we don’t use [Darmon et al. 1997, Lemma 4.11].

As a part of his proof, Kisin uses the description by Gee [2011] of Serre weights for \( \bar{\rho} \), which is available only for \( p > 2 \). We determine Serre weights for \( \bar{\rho} \) when \( p = 2 \) in Section 3D under assumption (iv) of Theorem 1.1. As in [Gee 2011] the main input is a modularity lifting theorem, which in our case is the theorem proved by Khare and Wintenberger [2009b] and Kisin [2009b]. We do this by a characteristic-0 argument, where Gee argues in characteristic \( p \); see Section 3D.

The modularity lifting theorems for \( p = 2 \) proved by Kisin [2009b], and more recently by Thorne [2016], do not require 2 to split completely in the totally real field \( F \), but they put a more restrictive hypothesis on \( \rho|_{\mathbb{G}_{F_v}} \) for \( v \mid 2 \). Kisin assumes that \( \rho|_{\mathbb{G}_{F_v}} \) for all \( v \mid 2 \) is potentially crystalline with Hodge–Tate weights equal to \((0,1)\) for every embedding \( F_v \hookrightarrow \overline{\mathbb{Q}}_2 \) and \( F_v = \mathbb{Q}_2 \) if \( \rho|_{\mathbb{G}_{F_v}} \) is ordinary. Thorne removes this last assumption, but requires instead that \( \bar{\rho}|_{\mathbb{G}_{F_v}} \) be nontrivial for at least one \( v \mid \infty \). We need 2 to split completely in \( F \) in order to apply the results on the \( p \)-adic Langlands correspondence, which is currently available only for GL\(_2(\mathbb{Q}_p)\).

2. Local part

2A. Capture. Let \( K \) be a profinite group with an open pro-\( p \) group. Let \( \mathcal{O}[[K]] \) be the completed group algebra, and let \( \text{Mod}_{K}^{\text{pro}}(\mathcal{O}) \) be the category of compact linear-topological \( \mathcal{O}[[K]] \)-modules. Let \( \psi : Z(K) \to \mathcal{O}^\times \) be a continuous character. We let \( \text{Mod}_{K,\psi}^{\text{pro}}(\mathcal{O}) \) be the full subcategory of \( \text{Mod}_{K}^{\text{pro}}(\mathcal{O}) \) such that \( M \in \text{Mod}_{K}^{\text{pro}}(\mathcal{O}) \) lies in \( \text{Mod}_{K,\psi}^{\text{pro}}(\mathcal{O}) \) if and only if \( Z(K) \) acts on \( M \) via \( \psi^{-1} \). Let \( \{V_i\}_{i \in I} \) be a family of continuous representations of \( K \) on finite-dimensional \( L \)-vector spaces, and let \( M \in \text{Mod}_{K}^{\text{pro}}(\mathcal{O}) \).

**Definition 2.1.** We say that \( \{V_i\}_{i \in I} \) captures \( M \) if the smallest quotient \( M \to Q \) such that \( \text{Hom}_{\mathcal{O}[[K]]}^{\text{cont}}(Q, V_i^*) \cong \text{Hom}_{\mathcal{O}[[K]]}^{\text{cont}}(M, V_i^*) \) for all \( i \in I \) is equal to \( M \).

We let \( c := \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} \) and note that the center of \( \text{SL}_2(\mathbb{Z}_p) \) is equal to \( \{1, c\} \).

**Lemma 2.2.** If \( K = \text{SL}_2(\mathbb{Z}_p) \) then \( \mathcal{O}[[K]]/(c-1) \) and \( \mathcal{O}[[K]]/(c+1) \) are \( \mathcal{O} \)-torsion-free.
Proof. If $K_n$ is an open normal subgroup of $K$ such that the image of $c$ in $K/K_n$ is nontrivial, then $O[K/K_n]$ is a free $O[Z]$-module, where $Z$ is the center of $K$. This implies that $O[K/K_n]/(c \pm 1)$ is a free $O$-module and by passing to the limit we obtain the assertion. □

Lemma 2.3. Let $K = \text{SL}_2(\mathbb{Z}_p)$, let $Z$ be the center of $K$ and let $\{V_i\}_{i \in I}$ be a family which captures $O[[K]]$ such that each $V_i$ has a central character. Let $I^+$ and $I^-$ be subsets of $I$ consisting of $i$ such that $c$ acts on $V_i$ by 1 and by $-1$, respectively. Let $\psi : Z \to L^\times$ be a character. If $\psi(c) = 1$ then $I^+$ captures every projective object in $\text{Mod}^\text{pro}_{K,\psi}(O)$. If $\psi(c) = -1$ then $I^-$ captures every projective object in $\text{Mod}^\text{pro}_{K,\psi}(O)$.

Proof. If $M \in \text{Mod}^\text{pro}_{K}(O)$ is $O$-torsion-free then $I$ captures $M$ if and only if the image of the evaluation map $\bigoplus_{i \in I} V_i \otimes \text{Hom}_K(V_i, \Pi) \to \Pi$ is dense, where $\Pi = \text{Hom}_{O}^\text{cont}(M, L)$ is the Banach space representation of $K$ with topology induced by the supremum norm [Colmez et al. 2014, Lemma 2.10]. Let $\Pi = \text{Hom}_{O}^\text{cont}(O[[K]], L)$ and $\Pi^\pm := \text{Hom}_{O}^\text{cont}(O[[K]]/(c \pm 1), L)$. Since $\Pi = \Pi^+ \oplus \Pi^-$, and $\{V_i\}$ captures $O[[K]]$, we deduce that the image of the evaluation map $\bigoplus_{i \in I} V_i \otimes \text{Hom}_K(V_i, \Pi^\pm) \to \Pi^\pm$ is dense. If $i \in I^+$ then $c$ acts trivially on $V_i$ and so $\text{Hom}_K(V_i, \Pi^-) = 0$. This implies the image of $\bigoplus_{i \in I^+} V_i \otimes \text{Hom}_K(V_i, \Pi^+) \to \Pi^+$ is dense. Using Lemma 2.2 we deduce that $I^+ \text{ captures } O[[K]]/(c-1)$. A similar argument shows that $I^- \text{ captures } O[[K]]/(c+1)$. Every projective object in $\text{Mod}^\text{pro}_{K,\psi}(O)$ can be realized as a direct summand of a product of some copies of $O[[K]]/(c - \psi(c))$, which implies the assertion; see the proof of [Colmez et al. 2014, Lemma 2.11]. □

Lemma 2.4. Let $K = \text{SL}_2(\mathbb{Z}_p)$, and let $Z$ be the center of $K$, $\psi : Z \to L^\times$ a character and $V$ a continuous representation of $K$ on a finite-dimensional $L$-vector space with a central character $\psi_V$. If $\psi(c) = \psi_V(c)$ then $\{V \otimes \text{Sym}^{2a} L^2\}_{a \in \mathbb{N}}$ captures every projective object in $\text{Mod}^\text{pro}_{K,\psi}(O)$; if $\psi(c) = -\psi_V(c)$ then $\{V \otimes \text{Sym}^{2a+1} L^2\}_{a \in \mathbb{N}}$ captures every projective object in $\text{Mod}^\text{pro}_{K,\psi}(O)$.

Proof. Proposition 2.12 in [Colmez et al. 2014] implies that $\{\text{Sym}^a L^2\}_{a \in \mathbb{N}}$ captures $O[[K]]$. We leave it as an exercise for the reader to check that this implies that $\{V \otimes \text{Sym}^a L^2\}_{a \in \mathbb{N}}$ also captures $O[[K]]$. The assertion follows from Lemma 2.3. □

Lemma 2.5. Let $M \in \text{Mod}^\text{pro}_{\text{GL}_2(\mathbb{Z}_p),\psi}(O)$ and let $V$ be a continuous representation of $K$ on a finite-dimensional $L$-vector space with a central character $\psi$. Then

$$\bigcap_{\phi} \text{Ker } \phi = \bigcap_{\xi,\eta} \text{Ker } \xi,$$

where the first intersection is taken over all $\phi \in \text{Hom}_{O[[\text{SL}_2(\mathbb{Z}_p)]]}^\text{cont}(M, V^*)$ and the second intersection is taken over all characters $\eta : \mathbb{Z}_p^\times \to L^\times$ with $\eta^2 = 1$ and all $\xi \in \text{Hom}_{O[[\text{GL}_2(\mathbb{Z}_p)]]}^\text{cont}(M, (V \otimes \eta \circ \text{det})^*)$.  

1308 Vytautas Paškūnas
Proof. Let $Z$ be the center of $GL_2(\mathbb{Z}_p)$. The determinant induces the isomorphism $GL_2(\mathbb{Z}_p)/Z \cong \mathbb{Z}_p^\times/(\mathbb{Z}_p^\times)^2$, which is a cyclic group of order 2 if $p \neq 2$, and a product of cyclic groups of order 2 if $p = 2$. Hence, $\text{Ind}_{Z SL_2(\mathbb{Z}_p)}^{GL_2(\mathbb{Z}_p)} 1 \cong \bigoplus \eta \circ \det$, where the sum is taken over all characters $\eta$ with $\eta^2 = 1$. The isomorphism
\[
\text{Hom}^\text{cont}_{O[SL_2(\mathbb{Z}_p)]}(M, V^*) \cong \text{Hom}^\text{cont}_{O[Z SL_2(\mathbb{Z}_p)]}(M, V^*) \\
\cong \text{Hom}^\text{cont}_{O[GL_2(\mathbb{Z}_p)]}(M, V^* \otimes \text{Ind}_{Z SL_2(\mathbb{Z}_p)}^{GL_2(\mathbb{Z}_p)} 1) \\
\cong \bigoplus_{\eta} \text{Hom}^\text{cont}_{O[GL_2(\mathbb{Z}_p)]}(M, V^* \otimes \eta \circ \det)
\]
is implied by the assertion. \hfill \square

Lemma 2.6. Let $M \in \text{Mod}^\text{pro}_{GL_2(\mathbb{Z}_p), \psi}(O)$ and let $\{V_i\}_{i \in I}$ be a family of continuous representations of $K$ on finite-dimensional $L$-vector spaces with a central character $\psi$. If $\{V_i|_{SL_2(\mathbb{Z}_p)}\}_{i \in I}$ captures $M|_{SL_2(\mathbb{Z}_p)}$ then $\{V_i \otimes \eta \circ \det\}_{i \in I, \eta}$ captures $M$, where $\eta$ runs over all characters $\eta : \mathbb{Z}_p^\times \rightarrow L^\times$ with $\eta^2 = 1$.

Proof. The assertion follows from Lemma 2.5 and [Colmez et al. 2014, Lemma 2.7]. \hfill \square

Proposition 2.7. Let $K = GL_2(\mathbb{Z}_p)$, and let $Z$ be the center of $K$ and $\psi : Z \rightarrow L^\times$ a continuous character. There is a smooth irreducible representation $\tau$ of $K$ which is a type for a Bernstein component containing a principal series representation, but not containing a special series representation, such that
\[
\{\tau \otimes \text{Sym}^a L^2 \otimes \eta' \circ \det\}_{a \in \mathbb{N}, \eta'}
\]
captures every projective object in $\text{Mod}^\text{pro}_{K, \psi}(O)$. Here, for each $a \in \mathbb{N}$, $\eta'$ runs over all continuous characters $\eta' : \mathbb{Z}_p^\times \rightarrow L^\times$ such that $\tau \otimes \text{Sym}^a L^2 \otimes \eta' \circ \det$ has central character $\psi$.

Proof. If $p \neq 2$ (resp. $p = 2$) then $1 + p\mathbb{Z}_p$ (resp. $1 + 4\mathbb{Z}_2$) is a free pro-$p$ group of rank 1. Using this one may show that there is a smooth, nontrivial character $\chi : \mathbb{Z}_p^\times \rightarrow L^\times$ and a continuous character $\eta_0 : \mathbb{Z}_p^\times \rightarrow L^\times$ such that $\psi = \chi \eta_0^2$. Let $e$ be the smallest integer such that $\chi$ is trivial on $1 + p^e\mathbb{Z}_p$. Let
\[
J = \left( \begin{array}{cc} \mathbb{Z}_p^\times & \mathbb{Z}_p^\times \\ p^e \mathbb{Z}_p & \mathbb{Z}_p^\times \end{array} \right),
\]
and let $\chi \otimes 1 : J \rightarrow L^\times$ be the character which sends $(a_{\chi, \eta_0}) \mapsto \chi(a)$. The representation $\tau := \text{Ind}^K_G(\chi \otimes 1)$ is irreducible and is a type. More precisely, for an irreducible smooth $\overline{L}$-representation $\pi$ of $G = GL_2(\mathbb{Q}_p)$, we have $\text{Hom}_K(\tau, \pi) \neq 0$ if and only if $\pi \cong (\text{Ind}^G_B \psi_1 \otimes \psi_2)_{\text{sm}}$, where $B$ is a Borel subgroup and $\psi_1|_{\mathbb{Z}_p^\times} = \chi$ and $\psi_2|_{\mathbb{Z}_p^\times} = 1$; see [Henniart 2002, §A.2.2]. The central character of $\tau$ is equal to $\chi$. We claim that the family $\{\tau \otimes \text{Sym}^{2a} L^2 \otimes (\det)^{-a} \otimes \eta \eta_0 \circ \det\}_{a \in \mathbb{N}, \eta}$, where $\eta$ runs
over all the characters with $\eta^2 = 1$, captures every projective object in $\text{Mod}_K^{\text{pro}}(O)$. If $M \in \text{Mod}_K^{\text{pro}}(O)$ is projective then $M|_{\text{SL}_2(\mathbb{Z}_p)}$ is projective in $\text{Mod}_\text{SL}_2(\mathbb{Z}_p, \psi)(O)$ [Emerton 2010b, Proposition 2.1.11]. Lemma 2.4 implies that the family captures $M|_{\text{SL}_2(\mathbb{Z}_p)}$. Since each representation in the family has central character equal to $\chi \eta_0^2 = \psi$, the claim follows from Lemma 2.6. Since the family of representations appearing in the claim is a subfamily of the representations appearing in the proposition, the claim implies the proposition. \hfill \Box

2B. The image of Colmez’s Montreal functor. Let $G = \text{GL}_2(\mathbb{Q}_p)$, $K = \text{GL}_2(\mathbb{Z}_p)$. Let $B$ be the subgroup of upper-triangular matrices in $G$, let $T$ be the subgroup of diagonal matrices and let $Z$ be the center of $G$. We make no assumption on the prime $p$. We fix a continuous character $\psi : Z \to O^\times$.

Let $\text{Mod}^{\text{pro}}_G(O)$ be the category of profinite augmented representations of $G$ [Emerton 2010a, Definition 2.1.6]. The Pontryagin duality

$$\pi \mapsto \pi^\vee := \text{Hom}_O^\text{cont}(\pi, L/O)$$

induces an antiequivalence of categories between $\text{Mod}_G^{\text{sm}}(O)$ and $\text{Mod}_G^{\text{pro}}(O)$ [Emerton 2010a, (2.2.8)]. Let $\text{Mod}_G^{\text{adm}}(O)$ be the full subcategory of $\text{Mod}_G^{\text{sm}}(O)$ consisting of locally admissible [Emerton 2010a, Definition 2.2.17] representations of $G$ and let $\text{Mod}_G^{\text{adm}}(O)$ be the full subcategory of $\text{Mod}_G^{\text{adm}}(O)$ consisting of those representations on which $Z$ acts by the character $\psi$. Let $\mathcal{C}(O)$ be the full subcategory of $\text{Mod}_G^{\text{pro}}(O)$ antiequivalent to $\text{Mod}_G^{\text{adm}}(O)$ via the Pontryagin duality. For $\pi_1, \pi_2 \in \text{Mod}_G^{\text{adm}}(O)$ we let $\text{Ext}^1_G(\pi_1, \pi_2)$ be the Yoneda Ext group computed in $\text{Mod}_G^{\text{adm}}(O)$.

Let $\pi \in \text{Mod}_G^{\text{adm}}(O)$ be absolutely irreducible and either supersingular [Barthel and Livné 1994; Breuil 2003a] or a principal series representation isomorphic to $\left(\text{Ind}_B^G \chi_1 \otimes \chi_2 \omega^{-1}\right)_{\text{sm}}$, for some smooth characters $\chi_1, \chi_2 : \mathbb{Q}_p^\times \to k^\times$ such that $\chi_1 \chi_2^{-1} \neq \omega^{\pm 1}, 1$. This hypothesis ensures that $\pi' := \left(\text{Ind}_B^G \chi_2 \otimes \chi_1 \omega^{-1}\right)_{\text{sm}}$ is also absolutely irreducible and $\pi \not\cong \pi'$. Let $P \to \pi^\vee$ be a projective envelope of $\pi^\vee$ in $\mathcal{C}(O)$ and let $E = \text{End}_{\mathcal{C}(O)}(P)$. Then $E$ is naturally a topological ring with a unique maximal ideal and residue field $k = \text{End}_{\mathcal{C}(O)}(\pi^\vee)$; see [Paškūnas 2013, §2].

Proposition 2.8. If $\pi$ is supersingular then $k \widehat{\otimes}_E P \cong \pi^\vee$. If $\pi$ is a principal series then $k \widehat{\otimes}_E P \cong \kappa^\vee$, where $\kappa$ is the unique nonsplit extension $0 \to \pi \to \kappa \to \pi' \to 0$.

Proof. In both cases, $(k \widehat{\otimes}_E P)^\vee$ is the unique representation in $\text{Mod}_G^{\text{adm}}(O)$ which is maximal with respect to the following conditions: (1) $\text{soc}(k \widehat{\otimes}_E P)^\vee \cong \pi$; (2) $\pi$ occurs in $(k \widehat{\otimes}_E P)^\vee$ with multiplicity one; see [Paškūnas 2013, Remark 1.13]. For, if $\tau \in \text{Mod}_G^{\text{adm}}(O)$ satisfies both conditions, then (1) and [Paškūnas 2013, Lemma 2.10] imply that the natural map $\text{Hom}_{\mathcal{C}(O)}(P, \tau^\vee) \widehat{\otimes}_E P \to \tau^\vee$ is surjective, and (2) and the exactness of $\text{Hom}_{\mathcal{C}(O)}(P, \ast)$ imply that $\text{Hom}_{\mathcal{C}(O)}(P, \tau^\vee) \cong \text{Hom}_{\mathcal{C}(O)}(P, \pi^\vee) \cong k$. Hence, dually we obtain an injection $\tau \leftrightarrow (k \widehat{\otimes}_E P)^\vee$. 

1310 Vytautas Paškūnas
Let $\pi_1$ be an irreducible representation in $\text{Mod}^{1, \text{adm}}_{G, \psi}(\mathcal{O})$ such that $\text{Ext}^1_{G, \psi}(\pi_1, \pi)$ is nonzero. It follows from Corollary 1.2 in [Paškūnas 2014] that if $\pi$ is supersingular then $\pi_1 \cong \pi$ and hence $(k \hat{\otimes}_E P)^{\vee} \cong \pi$, and if $\pi$ is a principal series as above then $\pi_1 \cong \pi$ or $\pi_1 \cong \pi'$. We will now explain how to modify the arguments of [Paškūnas 2013, §8] so that they also work for $p = 2$, the main point being that Emerton’s functor of ordinary parts works for all $p$. Proposition 4.3.15(2) of [Emerton 2010b] implies that $\text{Ext}^1_{G, \psi}(\pi', \pi)$ is one-dimensional. Let $\kappa$ be the unique nonsplit extension $0 \to \pi \to \kappa \to \pi' \to 0$. We claim that $\text{Ext}^n_{G, \psi}(\pi', \kappa) = 0$ for all $n \geq 0$. The claim for $n = 1$ implies that $(k \hat{\otimes}_E P)^{\vee} \cong \kappa$. It is proved in [Emerton and Paškūnas 2010, Corollary 3.12] that the $\delta$-functor $H \bullet \text{Ord}_B$, defined in [Emerton 2010b, Definition 3.3.1], is effaceable in $\text{Mod}^{1, \text{adm}}_{G, \psi}(\mathcal{O})$. Hence it coincides with the derived functor $R \bullet \text{Ord}_B$. An open compact subgroup $N_0$ of the unipotent radical of $B$ is isomorphic to $\mathbb{Z}_p$, and hence $H^i(N_0, \ast)$ vanishes for $i \geq 2$. This implies that $R^i \text{Ord}_B = H^i \text{Ord}_B = 0$ for $i \geq 2$. The proof of [Paškūnas 2013, Lemma 8.1] does not use the assumption $p > 2$ and gives that

$$\text{Ord}_B \kappa \cong \text{Ord}_B \pi \cong R^1 \text{Ord}_B \pi' \cong R^1 \text{Ord}_B \kappa \cong \chi_2 \omega^{-1} \otimes \chi_1.$$  \hspace{1cm} (2)

Our assumption on $\chi_1$ and $\chi_2$ implies that $\chi_1 \omega^{-1} \otimes \chi_2$ and $\chi_2 \omega^{-1} \otimes \chi_1$ are distinct characters of $T$. It follows from [Emerton 2010b, Lemma 4.3.10] that all the Ext-groups between them vanish. Since $\pi' \cong (\text{Ind}^G_B \chi_1 \omega^{-1} \otimes \chi_2)^{\text{sm}}$, where $\overline{B}$ is the subgroup of lower-triangular matrices in $G$, all the terms in Emerton’s spectral sequence [2010b, (3.7.4)] converging to $\text{Ext}^n_{G, \psi}(\pi', \kappa)$ are zero. Hence, $\text{Ext}^n_{G, \psi}(\pi_2, \kappa) = 0$ for all $n \geq 0$. Let us also note that the 5-term exact sequence associated to the spectral sequence implies that $\text{Ext}^1_{G, \psi}(\pi, \kappa)$ is finite-dimensional.

Proposition 2.9. If $\pi$ is supersingular then let $S = Q = \pi^{\vee}$. If $\pi$ is a principal series then let $S = \pi'$ and $Q = \kappa^{\vee}$. Then $S$ and $Q$ satisfy the hypotheses (H0)–(H5) of [Paškūnas 2013, §3].

Proof. If $\pi$ is supersingular then there are no other irreducible representations in the block of $\pi$ and hence the only hypothesis to check is (H4), which is equivalent to the finite-dimensionality of $\text{Ext}^1_{G, \psi}(\pi, \pi)$. This follows from Proposition 9.1 in [Paškūnas 2010b]. If $\pi$ is a principal series then the assertion follows from the Ext-group calculations made in the proof of Proposition 2.8.

The proposition enables us to apply the formalism developed in [Paškūnas 2013, Section 3]. Corollary 3.12 of [Paškūnas 2013] implies:

Proposition 2.10. The functor $\hat{\otimes}_E P$ is an exact functor from the category of pseudo-compact right $E$-modules to $\mathcal{C}(\mathcal{O})$. 
If $m$ is a pseudocompact right $E$-module then $\text{Hom}_{\mathcal{O}}(P, m \hat{\otimes}_E P) \cong m$ by [Paškūnas 2013, Lemma 2.9]. This implies that the functor is fully faithful, so that

$$\text{Hom}_{\mathcal{O}}^\text{cont}(m_1, m_2) \cong \text{Hom}_{\mathcal{O}}(m_1 \hat{\otimes}_E P, m_2 \hat{\otimes}_E P). \quad (3)$$

**Proposition 2.11.** $E$ is commutative.

*Proof.* Let $\tilde{\mathcal{C}}(\mathcal{O})$ be the full subcategory of $\text{Mod}_{\text{pro}}(\mathcal{O})$ which is antiequivalent to $\text{Mod}_{\text{ladn}}(\mathcal{O})$ via the Pontryagin duality. Let $\tilde{P}$ be a projective envelope of $\pi^\vee$ in $\tilde{\mathcal{C}}(\mathcal{O})$, let $\tilde{E} := \text{End}_{\tilde{\mathcal{C}}(\mathcal{O})}(\tilde{P})$ and let $a$ be the closed two-sided ideal of $\tilde{E}$ generated by the elements $z - \psi^{-1}(z)$, for all $z$ in the center of $G$. We may consider $\mathcal{C}(\mathcal{O})$ as a full subcategory of $\tilde{\mathcal{C}}(\mathcal{O})$. Since the center of $G$ acts on $\tilde{P}/a\tilde{P}$ by $\psi^{-1}$, we have $\tilde{P}/a\tilde{P} \in \mathcal{C}(\mathcal{O})$. The functor $\text{Hom}_{\mathcal{C}(\mathcal{O})}(\tilde{P}/a\tilde{P}, \ast)$ is exact, since

$$\text{Hom}_{\mathcal{C}(\mathcal{O})}(\tilde{P}/a\tilde{P}, M) = \text{Hom}_{\tilde{\mathcal{C}}(\mathcal{O})}(\tilde{P}, M) \quad (4)$$

for all $M \in \mathcal{C}(\mathcal{O})$, and $\tilde{P}$ is projective. Hence, $\tilde{P}/a\tilde{P}$ is projective in $\mathcal{C}(\mathcal{O})$. Its $G$-cosocle is isomorphic to $\pi^\vee$, since the same is true of $\tilde{P}$. Hence, $\tilde{P}/a\tilde{P}$ is a projective envelope of $\pi^\vee$ in $\mathcal{C}(\mathcal{O})$. Since projective envelopes are unique up to isomorphism, $\tilde{P}/a\tilde{P}$ is isomorphic to $P$. Since $a$ is generated by central elements, any $\phi \in \tilde{E}$ maps $a\tilde{P}$ to itself. This yields a ring homomorphism $\tilde{E} \to \text{End}_{\tilde{\mathcal{C}}(\mathcal{O})}(\tilde{P}/a\tilde{P}) \cong E$. Projectivity of $\tilde{P}$ and (4) applied with $M = \tilde{P}/a\tilde{P}$ implies that the homomorphism is surjective and induces an isomorphism $\tilde{E}/a \cong \text{End}_{\mathcal{C}(\mathcal{O})}(\tilde{P}/a\tilde{P})$. Since $\tilde{E}$ is commutative [Colmez et al. 2014, Corollary 2.22] we deduce that $E$ is commutative. □

**Proposition 2.12.** $E$ is a complete local noetherian commutative $\mathcal{O}$-algebra with residue field $k$.

*Proof.* Proposition 2.11 asserts that $E$ is commutative. Corollary 3.11 of [Paškūnas 2013] implies that the natural topology on $E$ (see [Paškūnas 2013, §2]) coincides with the topology defined by the maximal ideal $m$, which implies that $E$ is complete for the $m$-adic topology. It follows from Lemma 3.7, Proposition 3.8(iii) of [Paškūnas 2013] that $m/(m^2 + (\sigma))$ is a finite-dimensional $k$-vector space. Since $E$ is commutative, we deduce that $E$ is noetherian. □

**Proposition 2.13.** Let $Q = \pi^\vee$ if $\pi$ is supersingular and let $Q = \kappa^\vee$ if $\pi$ is a principal series. The ring $E$ represents the universal deformation problem of $Q$ in $\mathcal{C}(\mathcal{O})$, and $P$ is the universal deformation of $Q$.

*Proof.* Since $E$ is commutative by Proposition 2.11 and since hypotheses (H0)–(H5) of [Paškūnas 2013, §3] are satisfied by Proposition 2.9, the assertion follows from [Paškūnas 2013, Corollary 3.27]. □
2B1. Colmez’s Montreal functor. This subsection is essentially the same as Section 5.7 of [Paškūnas 2013]. Let $G_{\mathbb{Q}_p}$ be the absolute Galois group of $\mathbb{Q}_p$. We will consider $\psi$ as a character of $G_{\mathbb{Q}_p}$ via local class field theory, normalized so that the uniformizers correspond to geometric Frobenii. Let $\varepsilon : G_{\mathbb{Q}_p} \to \mathcal{O}_\mathbb{Q}^\times$ be the $p$-adic cyclotomic character. Similarly, we will identify $\varepsilon$ with the character of $\mathbb{Q}_p^\times$, which maps $x$ to $x/|x|$.

Colmez [2010] has defined an exact and covariant functor $V$ from the category of smooth, finite-length representations of $G$ on $\mathcal{O}$-torsion modules with a central character to the category of continuous finite-length representations of $G_{\mathbb{Q}_p}$ on $\mathcal{O}$-torsion modules. This functor enables us to make the connection between the $\text{GL}_2(\mathbb{Q}_p)$ and $G_{\mathbb{Q}_p}$ worlds. We modify Colmez’s functor to obtain an exact covariant functor $\tilde{V}$ as follows. Let $M$ be in $\mathcal{C}(\mathcal{O})$. If it is of finite length then $\tilde{V}(M) := V(M^\vee)^\vee(\varepsilon \psi)$, where $\vee$ denotes the Pontryagin dual and $\varepsilon$ is the cyclotomic character. In general, we may write $M \cong \varprojlim M_i$, where the limit is taken over all quotients of finite length in $\mathcal{C}(\mathcal{O})$, and we define $\tilde{V}(M) := \varprojlim \tilde{V}(M_i)$. If $\pi \in \text{Mod}_{G,\psi}(k)$ is absolutely irreducible, then $\pi^\vee$ is an object of $\mathcal{C}(\mathcal{O})$, and if $\pi$ is supersingular in the sense of [Barthel and Livné 1994], then $\tilde{V}((\pi^\vee) \cong V(\pi)$ is an absolutely irreducible continuous representation of $G_{\mathbb{Q}_p}$ associated to $\pi$ by Breuil [2003a].

The sequence is nonsplit by [Colmez 2010, Proposition VII.4.13(iii)]. If $m$ is a pseudocompact right $E$-module then there exists a natural isomorphism of $G_{\mathbb{Q}_p}$-representations

$$0 \to \chi_2 \to \tilde{V}(k^\vee) \to \chi_1 \to 0.$$  

The sequence is nonsplit by [Colmez 2010, Proposition VII.4.13(iii)]. If $m$ is a pseudocompact right $E$-module then there exists a natural isomorphism of $G_{\mathbb{Q}_p}$-representations

$$\tilde{V}(m \otimes_E P) \cong m \otimes_E \tilde{V}(P),$$  

by [Paškūnas 2013, Lemma 5.53]. It follows from (6) and Proposition 2.10 that $\tilde{V}(P)$ is a deformation of $\rho := \tilde{V}(k \otimes_E P)$ to $E$. If $\pi$ is supersingular then $\rho$ is an absolutely irreducible 2-dimensional representation of $G_{\mathbb{Q}_p}$, and if $\pi$ is a principal series then $\rho$ is a nonsplit extension of distinct characters; see (5). In both cases, $\text{End}_{G_{\mathbb{Q}_p}}(\rho) = k$ and so the universal deformation problem of $\rho$ is represented by a complete local noetherian $\mathcal{O}$-algebra $R$. Let $R^\psi$ be the quotient of $R$ parametrizing deformations of $\rho$ with determinant equal to $\psi \varepsilon$.

**Proposition 2.14.** The functor $\tilde{V}$ induces surjective homomorphisms $R \to E$ and $\varphi : E \to R^\psi$. 

Proof. This is proved in the same way as [Paškūnas 2013, Proposition 5.56, §5.8], following [Kisin 2010]. For the first surjection it is enough to prove that $\tilde{V}$ induces an injection

$$\text{Ext}^1_{\mathcal{O}(\mathcal{O})}(Q, Q) \hookrightarrow \text{Ext}^1_{G_{\mathbb{Q}_p}}(\rho, \rho).$$

This follows from [Colmez 2010, Théorème VII.5.2]. To prove the second surjection, we observe that $R^\psi$ is reduced and $\mathcal{O}$-torsion-free: if $p \geq 5$ then $R^\psi$ is formally smooth over $\mathcal{O}$, if $p = 3$ then the assertion follows from results of [Böckle 2010], and if $p = 2$ then the assertion follows from [Chenevier 2009, Proposition 4.1]. Thus it is enough to show that every closed point of $\text{Spec} R^2$ is contained in $\text{Spec} E$. This is equivalent to showing that for every deformation $\tilde{\rho}$ of $\rho$ with determinant $\psi \varepsilon$ there is a Banach space representation $\Pi$ lifting $Q$ with central character $\psi$ such that $\tilde{V}(\Pi) \cong \tilde{\rho}$. This follows from [Colmez et al. 2015, Theorem 10.1]. □

2B2. Banach space representations. Let $\text{Ban}_{G, \psi}^\text{adm}(L)$ be the category of admissible unitary $L$-Banach space representations [Schneider and Teitelbaum 2002, §3] on which $Z$ acts by the character $\psi$. If $\pi \in \text{Ban}_{G, \psi}^\text{adm}(L)$ then we let

$$\tilde{V}(\Pi) := \tilde{V}(\Theta^d) \otimes \mathcal{O} L,$$

where $\Theta$ is any open bounded $G$-invariant lattice in $\Pi$. Therefore, $\tilde{V}$ is exact and contravariant on $\text{Ban}_{G, \psi}^\text{adm}(L)$.

Remark 2.15. One of the reasons we use $\tilde{V}$ instead of $V$ is that this allows us to define $\tilde{V}(\Pi)$ without making the assumption that the reduction of $\Pi$ modulo $\varpi$ has finite length as a $G$-representation.

If $m$ is an $E[1/p]$-module of finite length then we let

$$\Pi(m) := \text{Hom}_{\mathcal{O}}^{\text{cont}}(m^0 \hat{\otimes}_E P, L),$$

where $m^0$ is any $E$-stable $\mathcal{O}$-lattice in $m$. Then $\Pi(m)$ is an admissible unitary $L$-Banach space representation of $G$, by [Paškūnas 2015b, Lemma 2.21], with the topology given by the supremum norm. Since the functor $\hat{\otimes}_E P$ is exact by Proposition 2.10, the functor $m \mapsto \Pi(m)$ is exact and contravariant. Moreover, it is fully faithful, as

$$\text{Hom}_G(\Pi(m_1), \Pi(m_2)) \cong \text{Hom}_{\mathcal{O}(\mathcal{O})}(m^0_1 \hat{\otimes}_E P, m^0_2 \hat{\otimes}_E P)_L$$

$$\cong \text{Hom}_{E[1/p]}(m_2, m_1),$$

where the first isomorphism follows from Theorem 2.3 of [Schneider and Teitelbaum 2002] and the second from (3).

Lemma 2.16. Let $m$ be an $E[1/p]$-module of finite length and let $\pi \in \text{Ban}_{G, \psi}^\text{adm}(L)$ be such that $\pi$ does not occur as a subquotient in the reduction of an open bounded
$G$-invariant lattice in $\Pi$ modulo $\varpi$. Then $\operatorname{Ext}_G^1(\Pi, \Pi(m))$ computed in $\operatorname{Ban}_{G, \psi}^{\text{adm}}(L)$ is zero.

Proof. If $\Theta$ is an open bounded $G$-invariant lattice in $B \in \operatorname{Ban}_{G, \psi}^{\text{adm}}(L)$ then we define $m(B) := \operatorname{Hom}_{\mathcal{O}}(P, \Theta^d)$. Proposition 4.17 in [Paškūnas 2013] implies that $m(B)$ is a finitely generated $E[1/p]$-module. The functor $B \mapsto m(B)$ is exact by [Paškūnas 2013, Lemma 4.9]. The evaluation map $\operatorname{Hom}_{\mathcal{O}}(P, \Theta^d) \otimes_E P \to \Theta^d$ induces a continuous $G$-equivariant map $B \to \Pi(m(B))$. If $m$ is an $E[1/p]$-module of finite length and $B \cong \Pi(m)$ then $m(B) \cong m$ and the map $B \to \Pi(m(B))$ is an isomorphism by [Paškūnas 2013, Lemma 4.28]. Moreover, $m(B) = 0$ if and only if $\pi$ does not occur as a subquotient of $\Theta/(\varpi)$, by [Colmez et al. 2014, Proposition 2.1(ii)]. Hence, if we have an exact sequence $0 \to \Pi(m) \to B \to \Pi \to 0$ then by applying the functor $m$ to it, we obtain an isomorphism $m \cong m(\Pi(m)) \cong m(B)$ and hence an isomorphism $\Pi(m) \cong \Pi(m(B))$. The map $B \to \Pi(m(B))$ splits the exact sequence. \qed

The proof of [Paškūnas 2015b, Lemma 4.3] shows that we have a natural isomorphism of $G_{\kappa(n)}$-representations

$$\tilde{V}(\Pi(m)) \cong m \otimes_E \tilde{V}(P). \quad (10)$$

Let us point out a special case of this isomorphism. If $n$ is a maximal ideal of $E[1/p]$ then its residue field $\kappa(n)$ is a finite extension of $L$. Let $\mathcal{O}_{\kappa(n)}$ be the ring of integers in $\kappa(n)$ and let $\varpi_{\kappa(n)}$ be the uniformizer. Then $\Theta := \operatorname{Hom}_{\mathcal{O}}^\text{cont}(\mathcal{O}_{\kappa(n)} \hat{\otimes}_E P, \mathcal{O})$ is an open bounded $G$-invariant lattice in $\Pi(\kappa(n))$. The evaluation map induces an isomorphism $\Theta^d \cong \mathcal{O}_{\kappa(n)} \hat{\otimes}_E P$. Since $E$ is noetherian, $\mathcal{O}_{\kappa(n)}$ is a finitely presented $E$-module and thus the usual and completed tensor products coincide. We obtain

$$\tilde{V}(\Theta^d) \cong \mathcal{O}_{\kappa(n)} \hat{\otimes}_E \tilde{V}(P), \quad \tilde{V}(\Pi(\kappa(n))) \cong \kappa(n) \hat{\otimes}_E \tilde{V}(P). \quad (11)$$

Since the residue field of $\mathcal{O}_{\kappa(n)}$ is $k$, we have

$$\Theta/(\varpi_{\kappa(n)}) \cong \operatorname{Hom}_k^\text{cont}(k \hat{\otimes}_E P, k) \cong (k \hat{\otimes}_E P)\hat{\vee}. \quad (12)$$

Recall from [Paškūnas 2013, §4] that $\Pi \in \operatorname{Ban}_{G, \psi}^{\text{adm}}(L)$ is irreducible if it does not have a nontrivial closed $G$-invariant subspace. It is absolutely irreducible if $\Pi \otimes_L L'$ is irreducible in $\operatorname{Ban}_{G, \psi}^{\text{adm}}(L')$ for every finite field extension $L'/L$. An irreducible $\Pi$ is ordinary if it is a subquotient of a unitary parabolic induction of a unitary character. Otherwise it is called nonordinary.

**Proposition 2.17.** If $n$ is a maximal ideal of $E[1/p]$ then either the $\kappa(n)$-Banach space representation $\Pi(\kappa(n))$ is absolutely irreducible nonordinary or

$$\pi \cong (\operatorname{Ind}_B^G \chi_1 \otimes \chi_2 \omega^{-1})_{\text{sm}}$$
and (after possibly replacing \( \kappa(n) \) by a finite extension) there exists a nonsplit extension
\[
0 \to \left( \text{Ind}_B^G \delta_1 \otimes \delta_2 e^{-1} \right)_{\text{cont}} \to \Pi(\kappa(n)) \to \left( \text{Ind}_B^G \delta_2 \otimes \delta_1 e^{-1} \right)_{\text{cont}} \to 0,
\]
where \( \delta_1, \delta_2 : \mathbb{Q}_p^\times \to \kappa(n)^\times \) are unitary characters congruent to \( \chi_1 \) and \( \chi_2 \), respectively, such that \( \delta_1 \delta_2 = \psi \varepsilon \).

**Proof.** It follows from (11) that \( \dim_{\kappa(n)} \check{\Phi}(\Pi(\kappa(n))) = 2 \). Since \( \check{\Phi} \) applied to a parabolic induction of a unitary character is a one-dimensional representation of \( G_{\mathbb{Q}_p} \), we deduce that if \( \Pi(\kappa(n)) \) is absolutely irreducible then it cannot be ordinary.

If \( \pi \) is supersingular then (12) implies that \( \Theta/((\sigma_{\kappa(n)})) \cong \pi \), which is absolutely irreducible. This implies that \( \Pi(\kappa(n)) \) is absolutely irreducible. If \( \pi \) is a principal series then \( \Theta/((\sigma_{\kappa(n)})) \) is of length 2 and both irreducible subquotients are absolutely irreducible. Hence, \( \Pi(\kappa(n)) \) is either irreducible or of length 2. Let us assume that \( \Pi(\kappa(n)) \) is not absolutely irreducible. Then after possibly replacing \( \kappa(n) \) by a finite extension we have an exact sequence of admissible \( \kappa(n) \)-Banach space representations
\[
0 \to \Pi_1 \to \Pi(\kappa(n)) \to \Pi_2 \to 0.
\]
This sequence is nonsplit, since otherwise \( \check{\Phi}(\Pi(\kappa(n))) \) would be a direct sum of two one-dimensional representations, which would contradict [Paškūnas 2015b, Lemma 4.5(iii)]. Let \( \Theta_1 := \Theta \cap \Pi_1 \) and let \( \Theta_2 \) be the image of \( \Theta \) in \( \Pi_2 \). Since we are dealing with admissible representations, \( \Theta_2 \) is a bounded \( \mathcal{O} \)-lattice in \( \Pi_2 \). Lemma 5.5 of [Paškūnas 2010a] says that we have the exact sequences of \( \mathcal{O}_{\kappa(n)} \)-modules
\[
0 \to \Theta_1 \to \Theta \to \Theta_2 \to 0,
\]
\[
0 \to \Theta_1/((\sigma_{\kappa(n)})) \to \Theta/((\sigma_{\kappa(n)})) \to \Theta_2/((\sigma_{\kappa(n)})) \to 0.
\]
It follows from (12) that the exact sequence of \( G \)-representations in (15) is the unique non-split extension \( 0 \to \pi \to \kappa \to \pi' \to 0 \). Proposition 4.2.14 of [Emerton 2010b] applied with \( A = \mathcal{O}_{\kappa(n)}/(\sigma_{\kappa(n)}^n) \) for all \( n \geq 1 \) implies that
\[
\Pi_1 \cong \left( \text{Ind}_B^G \delta_1 \otimes \delta_2 e^{-1} \right)_{\text{cont}}, \quad \Pi_2 \cong \left( \text{Ind}_B^G \delta'_1 \otimes \delta'_2 e^{-1} \right)_{\text{cont}},
\]
where \( \delta_1, \delta_2, \delta'_1, \delta'_2 : \mathbb{Q}_p^\times \to \kappa(n)^\times \) are unitary characters with \( \delta_1, \delta'_1 \) congruent to \( \chi_1 \) and \( \delta_2, \delta'_2 \) congruent to \( \chi_2 \) modulo \( \sigma_{\kappa(n)} \). We reduce (14) modulo \( \sigma_{\kappa(n)}^n \) to obtain an exact sequence to which we apply \( \text{Ord}_B \). This gives us an injection \( \text{Ord}_B(\Theta_2/((\sigma_{\kappa(n)}^n))) \hookrightarrow \mathbb{R}^1 \text{Ord}_B(\Theta_2/((\sigma_{\kappa(n)}^n))) \). Since both are free \( \mathcal{O}_{\kappa(n)}/(\sigma_{\kappa(n)}^n) \)-modules of rank 1, the injection is an isomorphism. This implies that \( \delta_1 \) is congruent to \( \delta'_1 \) and \( \delta_2 \) is congruent \( \delta'_2 \) modulo \( \sigma_{\kappa(n)}^n \) for all \( n \geq 1 \). Hence, \( \delta_1 = \delta'_1 \) and \( \delta_2 = \delta'_2 \). \( \square \)

**2B3. Main local result.** We will prove that the surjection \( \varphi : E \to R^\psi \) in Proposition 2.14 is an isomorphism. The argument combines the first part of the paper with methods of [Paškūnas 2015b]. The argument in [Paškūnas 2013] used to prove this statement when \( p \geq 5 \) uses the fact that the rings \( R^\psi \) are formally smooth in that
case. This does not hold in general; when $p = 2$ or $3$ and even when the ring is formally smooth and $p = 3$, the computations just get too complicated.

Let $V$ be a continuous representation of $K$ with a central character $\psi$ of the form $\tau \otimes \text{Sym}^a L^2 \otimes \eta \circ \det$, where $\eta : \mathbb{Z}_p^\times \rightarrow L^\times$ is a continuous character, and $\tau$ is a type for a Bernstein component containing a principal series representation, but not containing a special series representation.

**Proposition 2.18.** If $n$ is a maximal ideal of $E[1/p]$ then the following hold:

(i) $\dim_{\kappa(n)} \text{Hom}_K(V, \Pi(\kappa(n))) \leq 1$.

(ii) $\dim_{\kappa(n)} \text{Hom}_K(V, \Pi(E_n/n^2)) \leq 2$.

Moreover, if $\text{Hom}_K(V, \Pi(\kappa(n))) \neq 0$ then $\det \tilde{V}(\Pi(\kappa(n))) = \psi \varepsilon$.

**Proof.** If $m$ is an $E[1/p]$-module of finite length and $L'$ is a finite extension of $L$, then $\Pi(m \otimes_L L') \cong \Pi(m) \otimes_L L'$ and $\text{Hom}_K(V, \Pi(m)) \otimes_L L' \cong \text{Hom}_K(V, \Pi(m \otimes_L L'))$. This implies that it is enough to prove the assertions after replacing $\kappa(n)$ by a finite extension. In particular, we may assume that $\Pi(\kappa(n))$ is either absolutely irreducible or a nonsplit extension as in Proposition 2.17. Since $\tilde{V}$ is compatible with twisting by characters, to prove the proposition it is enough to assume that $\eta$ is trivial, so that $V$ is a locally algebraic representation of $K$.

Since $\tau$ is a type and $\Pi(\kappa(n))$ is admissible, $\text{Hom}_K(V, \Pi(\kappa(n))) \neq 0$ if and only if (after possibly replacing $\kappa(n)$ by a finite extension) $\Pi(\kappa(n))$ contains a subrepresentation of the form $\Psi \otimes \text{Sym}^a L^2$, where $\Psi$ is an absolutely irreducible smooth principal series representation in the Bernstein component described by $\tau$; see the proof of [Paškūnas 2010a, Theorem 7.2]. Let $\Pi$ be the universal unitary completion of $\Psi \otimes \text{Sym}^a L^2$. Then $\Pi$ is absolutely irreducible, by [Berger and Breuil 2010, Corollaire 5.3.4] and [Breuil and Emerton 2010, Proposition 2.2.1].

If $\Pi(\kappa(n))$ is absolutely irreducible, we deduce that $\Pi(\kappa(n)) \cong \Pi$. Since $\Pi$ in [Berger and Breuil 2010] is constructed out of a $(\varphi, \Gamma)$-module of a 2-dimensional crystabeline representation of $G_{\mathbb{Q}_p}$ with determinant $\psi \varepsilon$, applying $\tilde{V}$ undoes this construction to obtain the Galois representation we started with. In particular, $\det \tilde{V}(\Pi(\kappa(n))) = \psi \varepsilon$. Moreover, it follows from [Colmez 2010, Théorème VI.6.50] that the locally algebraic vectors in $\Pi(\kappa(n))$ are isomorphic to $\Psi \otimes \text{Sym}^a L^2$, which implies that

$$\dim_{\kappa(n)} \text{Hom}_K(V, \Pi(\kappa(n))) = \dim_{\kappa(n)} \text{Hom}_K(V, \Psi \otimes \text{Sym}^a L^2) = 1,$$

(16)

giving part (i).

If $\Pi(\kappa(n))$ is reducible, then using the fact that (13) is nonsplit we deduce that $\Pi$ is the unique irreducible subrepresentation of $\Pi(\kappa(n))$. It follows from [Paškūnas
where \( d \) of finite length then appealing to Theorem 11.4. All the other arguments in that section work for all primes \( p \) lies in the image of \( m\text{-Spec} \).

It follows from [Paškūnas 2015b, Lemma 2.33] that \( \psi \) is \( O \)-torsion-free. Moreover \( \psi \) is a finitely generated \( E \)-module. Hence, [Paškūnas 2015b, Proposition 2.15] implies that \( \text{Ext}_G^1(\Pi'(\kappa(n)),5(\kappa(n))) = 0. \) Hence, \( \text{Hom}_G(\Pi'(\kappa(n)),(E_n/n^2)) \cong \text{Hom}_G(\Pi(\kappa(n)),(E_n/n^2)) \) and the claim follows from (9).

Let \( \Theta \) be a \( K \)-invariant \( O \)-lattice in \( V \) and let \( M(\Theta) := \text{Hom}_{O[\kappa]}(P, \Theta^d)^d \), where \((*)^d := \text{Hom}_O(*, O). \) It follows from Proposition 2.8 that \( (k \hat{\otimes}_E P)^\vee \) is an admissible representation of \( G; \) dually, this implies that \( k \hat{\otimes}_E P \) is a finitely generated \( O[\kappa] \)-module. Hence, [Paškūnas 2015b, Proposition 2.15] implies that \( M(\Theta) \) is a finitely generated \( E \)-module. We will denote by \( m\text{-Spec} \) the set of maximal ideals of a commutative ring.

**Proposition 2.19.** Let \( a \) be the \( E \)-annihilator of \( M(\Theta) \). Then \( E/a \) is reduced and \( O \)-torsion-free. Moreover, \( m\text{-Spec}(E/a)[1/p] \) is contained in the image of \( m\text{-Spec} R^\psi[1/p] \) under \( \varphi^a : \text{Spec} R^\psi \to \text{Spec} E. \)

**Proof.** Theorem 5.2 in [Paškūnas 2015b] implies that there is a \( P \)-regular \( x \in E \) such that \( P/xP \) is a finitely generated \( O[\kappa] \)-module which is projective in \( \text{Mod}_{\text{pro}}^\psi(\mathcal{O}). \) It follows from [Paškūnas 2015b, Lemma 2.33] that \( M(\Theta) \) is Cohen–Macaulay as a module over \( E \) and its Krull dimension is equal to 2. If \( m \) is an \( E[1/p] \)-module of finite length then

\[
\text{dim}_L \text{Hom}_K(V, \Pi(m)) = \text{dim}_L m \otimes E M(\Theta),
\]

by [Paškūnas 2015b, Proposition 2.22]. Proposition 2.18 together with [Paškūnas 2015b, Proposition 2.32] imply that \( E/a \) is reduced. It is \( O \)-torsion-free, since \( M(\Theta) \) is \( O \)-torsion-free. Let \( n \) be a maximal ideal of \( E[1/p] \). Since \( E \) is a quotient of \( R \), \( n \) lies in the image of \( m\text{-Spec} R^\psi[1/p] \) if and only if \( \det \kappa(n) \otimes E \hat{V}(P) = \psi \).

\[1\] The assumption \( p \geq 5 \) in [Paškūnas 2013, §12] is only invoked in the proof of Theorem 12.7 by appealing to Theorem 11.4. All the other arguments in that section work for all primes \( p \).
Proposition 2.18, (11) and (17) imply that this holds for all the maximal ideals of $(E/\mathfrak{a})[1/p]$. □

**Corollary 2.20.** The surjection $\varphi : E \to R^\psi$, given by Proposition 2.14, induces an isomorphism $E/\mathfrak{a} \cong R^\psi/\varphi(\mathfrak{a})$.

**Proof.** Since $(E/\mathfrak{a})[1/p]$ and $(R^\psi/\varphi(\mathfrak{a}))[1/p]$ are Jacobson, Proposition 2.19 implies that $\varphi$ induces an isomorphism between $E/\mathfrak{a}$ and the image of $R^\psi$ in the maximal reduced quotient of $(R^\psi/\varphi(\mathfrak{a}))[1/p]$. This implies that the surjection $E/\mathfrak{a} \to R^\psi/\varphi(\mathfrak{a})$ is injective, and hence an isomorphism. □

**Lemma 2.21.** The $E$-annihilators of $\text{Hom}^\text{cont}_K(P, V^*)$ and $M(2)$ are equal.

**Proof.** One inclusion is trivial; the other follows from [Paškūnas 2015b, (11)], which says that $\text{Hom}^\text{cont}_K(P, V^*)$ is naturally isomorphic to $\text{Hom}^\text{cont}_O(M(2), L)$. □

**Theorem 2.22.** The functor $\check{V}$ induces an isomorphism $\varphi : E \to R^\psi$. Moreover, $\check{V}(P)$ is the universal deformation of $\rho$ with determinant $\psi\varepsilon$.

**Proof.** It follows from Corollary 2.20 and Lemma 2.21 that the kernel of $\varphi$ is contained in the $E$-annihilator of $\text{Hom}^\text{cont}_K(P, V^*)$. It follows from Proposition 2.7 that the intersection of the annihilators as $V$ varies is zero. Hence, $\varphi$ is injective, and hence an isomorphism by Proposition 2.14. The second part is a formal consequence of the first part. □

**2B4. Blocks.** As explained in the introduction the category $\text{Mod}_{G, \psi}^\text{adm}(\mathcal{O})$ decomposes into a product of subcategories

$$\text{Mod}_{G, \psi}^\text{adm}(\mathcal{O}) \cong \prod_{\mathfrak{B} \in \text{Irr}_{G, \psi}^\text{adm} / \sim} \text{Mod}_{G, \psi}^\text{adm}(\mathcal{O})[\mathfrak{B}],$$

where $\text{Mod}_{G, \psi}^\text{adm}(\mathcal{O})[\mathfrak{B}]$ is the full subcategory of $\text{Mod}_{G, \psi}^\text{adm}(\mathcal{O})$ consisting of representations with all irreducible subquotients in $\mathfrak{B}$. Dually we obtain a decomposition

$$\mathfrak{C}(\mathcal{O}) \cong \prod_{\mathfrak{B} \in \text{Irr}_{G, \psi}^\text{adm} / \sim} \mathfrak{C}(\mathcal{O})[\mathfrak{B}],
$$

where $M \in \mathfrak{C}(\mathcal{O})$ lies in $\mathfrak{C}(\mathcal{O})[\mathfrak{B}]$ if and only if $M^\vee$ lies in $\text{Mod}_{G, \psi}^\text{adm}(\mathcal{O})[\mathfrak{B}]$.

For a block $\mathfrak{B}$ let $\pi_{\mathfrak{B}} = \bigoplus_{\pi \in \mathfrak{B}} \pi$, and let $\pi_{\mathfrak{B}} \hookrightarrow J_{\mathfrak{B}}$ be an injective envelope of $\pi_{\mathfrak{B}}$. Then $P_{\mathfrak{B}} := (J_{\mathfrak{B}})^\vee$ is a projective envelope of $(\pi_{\mathfrak{B}})^\vee$ in $\mathfrak{C}(\mathcal{O})$. Moreover, $J_{\mathfrak{B}}$ is an injective generator of $\text{Mod}_{G, \psi}^\text{adm}(\mathcal{O})[\mathfrak{B}]$ and $P_{\mathfrak{B}}$ is a projective generator of $\mathfrak{C}(\mathcal{O})[\mathfrak{B}]$. The ring $E_{\mathfrak{B}} := \text{End}_{\mathfrak{C}(\mathcal{O})}(P_{\mathfrak{B}})$ carries a natural topology with respect to which it is a pseudocompact ring; see [Gabriel 1962, Chapitre IV, Proposition 13]. In addition, the functor

$$M \mapsto \text{Hom}_{\mathfrak{C}(\mathcal{O})}(P_{\mathfrak{B}}, M)$$
induces an equivalence of categories between $\mathcal{C}(O)[B]$ and the category of right pseudocompact $E_B$-modules; see Corollaire 1 after [Gabriel 1962, Chapitre IV, Théorème 4]. The inverse functor is given by $m \mapsto m \hat{\otimes}_E P_B$, as follows from Lemmas 2.9 and 2.10 in [Paškūnas 2013]. Moreover, the center of the category of $\mathcal{C}(O)[B]$, which by definition is the ring of the natural transformations of the identity functor, is naturally isomorphic to the center of the ring $E_B$; see Corollaire 5 after [Gabriel 1962, Chapitre IV, Théorème 4].

Let us prove Theorem 1.2, stated in the introduction. If $B$ is a block containing a supersingular representation $\pi$ then $B = \{\pi\}$ and so $\pi_B = \pi$, $P_B$ is a projective envelope of $\pi^\vee$ and $E_B$ coincides with the ring denoted by $E$ in the previous section. Theorem 2.22 implies that $E_B$ is naturally isomorphic to $R_\psi^\rho$, the quotient of the universal deformation ring of $\rho := \hat{\nabla}(\pi^\vee)$ parametrizing deformations with determinant $\psi \varepsilon$. Since this ring is commutative, we deduce that the center of $\mathcal{C}(O)[B]$ is naturally isomorphic to the center of $R_\psi^\rho$. Moreover, $\hat{\nabla}(P_B)$ is the tautological deformation of $\rho$ to $R_\psi^\rho$; see Theorem 2.22.

If $B$ contains a generic principal series representation then $B = \{\pi_1, \pi_2\}$, where

$$\pi_1 \cong (\text{Ind}^G_B \chi_1 \otimes \chi_2 \omega^{-1})_{\text{sm}}, \quad \pi_2 \cong (\text{Ind}^G_B \chi_2 \otimes \chi_1 \omega^{-1})_{\text{sm}},$$

and $\chi_1, \chi_2 : \mathbb{Q}_p^\times \to k^\times$ are continuous characters such that $\chi_1 \chi_2^{-1} \neq 1, \omega^{\pm 1}$. Then $\pi_B = \pi_1 \oplus \pi_2$ and so $P_B \cong P_1 \oplus P_2$, where $P_1$ is a projective envelope of $\pi_1^\vee$ and $P_2$ is a projective envelope of $\pi_2^\vee$ in $\mathcal{C}(O)$. Thus

$$E_B \cong \text{End}_{\mathcal{C}(O)}(P_1 \oplus P_2) \cong \text{End}_{\text{cont}}^{\text{G}_{\mathbb{Q}_p}}(\hat{\nabla}(P_1) \oplus \hat{\nabla}(P_2)),$$

where the last isomorphism follows from [Paškūnas 2013, Lemma 8.10]. The assumption on the characters $\chi_1, \chi_2$ implies that if we consider them as representations of $G_{\mathbb{Q}_p}$ via the local class field theory, $\text{Ext}^1$-groups between them are 1-dimensional. This means there are unique up to isomorphism nonsplit extensions

$$\rho_1 = \begin{pmatrix} \chi_1 & * \\ 0 & \chi_2 \end{pmatrix}, \quad \rho_2 = \begin{pmatrix} \chi_1 & 0 \\ * & \chi_2 \end{pmatrix}.$$

Let $R_1$ be the universal deformation ring of $\rho_1$, let $R_1^\rho_\psi$ be the quotient of $R_1$ parametrizing deformations of $\rho_1$ with determinant $\psi \varepsilon$, and let $\rho_1^{\text{univ}}$ be the tautological deformation of $\rho_1$ to $R_1^\rho_\psi$. We define $R_2, R_2^\rho_\psi$ and $\rho_2^{\text{univ}}$ in the same way with $\rho_2$ instead of $\rho_1$. It follows from Theorem 2.22 and (21) that

$$E_B \cong \text{End}_{\text{cont}}^{\text{G}_{\mathbb{Q}_p}}(\rho_1^{\text{univ}} \oplus \rho_2^{\text{univ}}).$$

We have studied the right-hand side of (22) in [Paškūnas 2013, §B.1] for $p > 2$ and in [Paškūnas 2015a] in general. To describe the result we need to recall the theory of determinants due to Chenevier [2014].
Let $\rho : G_{\mathbb{Q}_p} \to \text{GL}_2(k)$ be a continuous representation. Let $\mathfrak{A}$ be the category of local artinian augmented $\mathcal{O}$-algebras with residue field $k$. Let $D^{ps} : \mathfrak{A} \to \text{Sets}$ be the functor which maps $(A, m_A) \in \mathfrak{A}$ to the set of pairs of functions $(t, d) : G_{\mathbb{Q}_p} \to A$ such that:

- $d : G_{\mathbb{Q}_p} \to A^\times$ is a continuous group homomorphism, congruent to $\det \rho$ modulo $m_A$.
- $t : G_{\mathbb{Q}_p} \to A$ is a continuous function with $t(1) = 2$.
- For all $g, h \in G_{\mathbb{Q}_p}$, the following are satisfied:
  
  (i) $t(g) \equiv \text{tr} \rho(g) \pmod{m_A}$.
  (ii) $t(gh) = t(hg)$.
  (iii) $d(g)t(g^{-1}h) - t(g)t(h) + t(gh) = 0$.

The functor $D^{ps}$ is prorepresented by a complete local noetherian $\mathcal{O}$-algebra $R^{ps}$. Let $R^{ps, \psi}$ be the quotient of $R^{ps}$ parametrizing those pairs $(t, d)$ where $d = \psi \epsilon$. Combining (22) with [Paškūnas 2015a, Propositions 3.12 and 4.3, Corollary 4.4] we obtain the following:

**Theorem 2.23.** Let $\mathcal{B} = \{\pi_1, \pi_2\}$ as above and let $\rho = \chi_1 \oplus \chi_2$. The center of $E_{\mathcal{B}}$, and hence the center of the category $\mathcal{E}(\mathcal{O})[\mathcal{B}]$, is naturally isomorphic to $R^{ps, \psi}$. Moreover, $E_{\mathcal{B}}$ is a free $R^{ps, \psi}$-module of rank 4:

$$E_{\mathcal{B}} \cong \left( R^{ps, \psi} e_{\chi_1}, R^{ps, \psi} \tilde{\Phi}_{12}, R^{ps, \psi} e_{\chi_2}, R^{ps, \psi} \Phi_{21} \right).$$

The generators satisfy the following relations:

$$e_{\chi_1}^2 = e_{\chi_1}, \quad e_{\chi_2}^2 = e_{\chi_2}, \quad e_{\chi_1} e_{\chi_2} = e_{\chi_2} e_{\chi_1} = 0,$$

$$e_{\chi_1} \tilde{\Phi}_{12} = \tilde{\Phi}_{12} e_{\chi_2} = \tilde{\Phi}_{12}, \quad e_{\chi_2} \tilde{\Phi}_{21} = \tilde{\Phi}_{21} e_{\chi_1} = \tilde{\Phi}_{21},$$

$$e_{\chi_2} \tilde{\Phi}_{12} = \tilde{\Phi}_{12} e_{\chi_1} = e_{\chi_1} \tilde{\Phi}_{21} = \tilde{\Phi}_{21} e_{\chi_2} = \tilde{\Phi}_{12}^2 = \tilde{\Phi}_{21}^2 = 0,$$

$$\tilde{\Phi}_{12} \tilde{\Phi}_{21} = c e_{\chi_1}, \quad \tilde{\Phi}_{21} \tilde{\Phi}_{12} = c e_{\chi_2}. $$

The element $c$ is regular in $R^{ps, \psi}$ and generates the reducibility ideal.

In order to state the result about the center of $\mathcal{E}(\mathcal{O})[\mathcal{B}]$ in a uniform way, as in Theorem 1.3, we note that if $\rho$ is an irreducible representation then mapping a deformation $\rho_A$ to $(\text{tr} \rho_A, \det \rho_A)$ induces a homomorphism of $\mathcal{O}$-algebras $R^{ps} \to R_\rho$, which is an isomorphism by [Chenevier 2014, Theorem 2.22, Example 3.4].

For a block $\mathcal{B}$, let $\text{Ban}_{G, \psi}(L)[\mathcal{B}]$ be the full subcategory of $\text{Ban}_{G, \psi}(L)$ consisting of those $\Pi$ for which, for some (equivalently any) open bounded $G$-invariant lattice $\Theta$, all the irreducible subquotients of $\Theta \otimes_{\mathcal{O}} k$ lie in $\mathcal{B}$. It is shown in
[Paškūnas 2013, Proposition 5.36] that $\text{Ban}^\text{adm}_{G,\psi}(L)$ decomposes into a direct sum of subcategories

$$\text{Ban}^\text{adm}_{G,\psi}(L) \cong \bigoplus_{\mathfrak{B} \in \text{Irr}^\text{adm}_{G}/\sim} \text{Ban}^\text{adm}_{G,\psi}(L)[\mathfrak{B}].$$

**Corollary 2.24.** If $\mathfrak{B} = \{\pi\}$ with $\pi$ supersingular then let $\rho = \tilde{V}(\pi^\vee)$. If $\mathfrak{B} = \{\pi_1, \pi_2\}$ with $\pi_1, \pi_2$ given by (20) then let $\rho = \tilde{V}(\pi_1^\vee) \oplus \tilde{V}(\pi_2^\vee) = \chi_1 \oplus \chi_2$. The map $\Pi \mapsto \tilde{V}(\Pi)$ induces a bijection between the isomorphism classes of

- absolutely irreducible nonordinary $\Pi \in \text{Ban}^\text{adm}_{G,\psi}(L)[\mathfrak{B}]$;
- absolutely irreducible $\tilde{\rho} : G_{Q_p} \to \text{GL}_2(L)$ such that $\det \tilde{\rho} = \psi \varepsilon$ and the semisimplification of the reduction modulo $\varpi$ of a $G_{Q_p}$-invariant $O$-lattice in $\tilde{\rho}$ is isomorphic to $\rho$.

**Proof.** Given Theorems 1.2 and 2.23, this is proved in the same way as [Paškūnas 2013, Theorem 11.4].

If $\Pi \in \text{Ban}^\text{adm}_{G,\psi}(L)[\mathfrak{B}]$ and $\Theta$ is an open bounded $G$-invariant lattice in $\Pi$, then $\Theta / \varpi^n$ is an object of $\text{Mod}^\text{adm}_{G,\psi}(O)[\mathfrak{B}]$ for all $n \geq 1$. Theorem 1.3 gives a natural action of $R^\text{ps,} \psi$ on $\Theta / \varpi^n$ for all $n \geq 1$. Passing to the limit and inverting $p$, we obtain a natural homomorphism $R^\text{ps,} \psi [1/p] \to \text{End}^\text{cont}_G(\Pi)$.

**Corollary 2.25.** Let $\mathfrak{B}$ be as in Corollary 2.24 and let $\Pi \in \text{Ban}^\text{adm}_{G,\psi}(L)[\mathfrak{B}]$ be absolutely irreducible. Then $\text{tr} \tilde{V}(\Pi)$ is equal to the specialization of the universal pseudocharacter $t^{\text{univ}} : G_{Q_p} \to R^\text{ps,} \psi$ at $x : R^\text{ps,} \psi \to \text{End}^\text{cont}_G(\Pi) \cong L$.

**Proof.** This is proved in the same way as [Paškūnas 2013, Proposition 11.3]. To carry out that proof we need to verify that $\tilde{V}(P_{\mathfrak{B}})$ is annihilated by $g^2 - t^{\text{univ}}(g)g + \psi \varepsilon(g)$ for all $g \in G_{Q_p}$. If $\mathfrak{B}$ contains a supersingular representation this follows from Cayley–Hamilton since $\tilde{V}(P_{\mathfrak{B}})$ is the universal deformation of $\rho$ with determinant $\psi \varepsilon$, and $\text{tr} \tilde{V}(P_{\mathfrak{B}}) = t^{\text{univ}}$ by [Chenevier 2014, Theorem 2.22, Example 3.4]. If $\mathfrak{B}$ contains a generic principal series then $\tilde{V}(P_{\mathfrak{B}}) \cong \rho_1^{\text{univ}} \oplus \rho_2^{\text{univ}}$ and the assertion follows from [Paškūnas 2015a, Proposition 3.9].

**Corollary 2.26.** For any $\Pi$ as in Corollary 2.24, we have $\dim_L \text{Ext}^1_{G,\psi}(\Pi, \Pi) = 3$.

**Proof.** Let $\text{Ban}^\text{adm,fl}_{G,\psi}(L)[\mathfrak{B}]$ be the full subcategory of $\text{Ban}^\text{adm}_{G,\psi}(L)[\mathfrak{B}]$ consisting of objects of finite length. It follows from [Paškūnas 2013, Theorem 4.36] that this category decomposes into a direct sum of subcategories

$$\text{Ban}^\text{adm,fl}_{G,\psi}(L)[\mathfrak{B}] \cong \bigoplus_{n \in \text{m-Spec} R^\text{ps,} \psi[1/p]} \text{Ban}^\text{adm,fl}_{G,\psi}(L)[\mathfrak{B}]_n,$$

where, for a maximal ideal $n$ of $R^\text{ps,} \psi[1/p]$, the direct summand $\text{Ban}^\text{adm,fl}_{G,\psi}(L)[\mathfrak{B}]_n$ consists of those finite-length representations which are killed by a power of $n$. Moreover, the last part of [Paškūnas 2013, Theorem 4.36] implies that the functor
$\Pi \mapsto \text{Hom}_{\mathcal{O}}(P_{\mathbb{B}}, \Theta^{d})[1/p]$, where $\Theta$ is any open bounded $G$-invariant lattice in $\Pi$, induces an antiequivalence of categories between $\text{Ban}_{G, \psi}^{\text{adm, fl}}(L)[\mathbb{B}]_{n}$ and the category of modules of finite length over the $n$-adic completion of $E_{\mathbb{B}}[1/p]$, which we denote by $\widehat{E}_{\mathbb{B}, n}$.

Let $\tilde{\rho} = \check{V}(\Pi)$. Corollary 2.24 tells us that $\tilde{\rho}$ is an absolutely irreducible representation with determinant $\psi \varepsilon$. Let $n$ be the maximal ideal of $R^{\text{ps}, \psi}[1/p]$ corresponding to the pair $(\text{tr} \tilde{\rho}, \text{det} \tilde{\rho})$. It follows from Corollary 2.25 that $\Pi$ is annihilated by $n$ and hence lies in $\text{Ban}_{G, \psi}^{\text{adm, fl}}(L)[\mathbb{B}]_{n}$. Let $A$ be the completion of $R^{\text{ps}, \psi}[1/p]$ at $n$. In the supersingular case, $E_{\mathbb{B}} = R^{\text{ps}, \psi} = R^{\psi}$, and so $\widehat{E}_{\mathbb{B}, n} = A$. In the generic principal series case, since $\tilde{\rho}$ is absolutely irreducible, the image of the generator of the reducible locus in $R^{\text{ps}, \psi}$ in $\kappa(n)$ is nonzero. It follows from the description of $E_{\mathbb{B}}$ in Theorem 2.23 that $\widehat{E}_{\mathbb{B}, n}$ is isomorphic to the algebra of $2 \times 2$ matrices with entries in $A$. Thus in both cases we get that $\text{Ban}_{G, \psi}^{\text{adm, fl}}(L)[\mathbb{B}]_{n}$ is antiequivalent to the category of $A$-modules of finite length, and $\Pi$ is identified with the residue field $\kappa(n)$ of $A$. Hence,

$$\text{Ext}_{G, \psi}^{1}(\Pi, \Pi) \cong \text{Ext}_{A}^{1}(\kappa(n), \kappa(n)).$$

Arguing as in [Kisin 2009c, Lemma 2.3.3] we may identify $A$ with the universal deformation ring parameterizing pseudocharacters with determinant $\psi \varepsilon$ and values in local artinian $L$-algebras which lift $\text{tr} \tilde{\rho}$. Since $\tilde{\rho}$ is absolutely irreducible we may further identify this ring with the quotient of the universal deformation ring of $\tilde{\rho}$ to local artinian $L$-algebras parameterizing deformations with determinant $\psi \varepsilon$. This ring is formally smooth over $L$ of dimension $3$, as $H^{2}(G_{\mathbb{Q}_{p}}, \text{ad}^{0}(\tilde{\rho})) \cong H^{0}(G_{\mathbb{Q}_{p}}, \text{ad}^{0}(\tilde{\rho})(1)) = 0$ and so the deformation problem of $\tilde{\rho}$ is unobstructed. In particular, $\dim_{L} \text{Ext}_{A}^{1}(\kappa(n), \kappa(n)) = \dim_{L} nA/n^{2}A = 3$. \hfill $\square$

2C. The Breuil–Mézard conjecture. In this section we apply the formalism developed in [Paškūnas 2015b] to prove new cases of the Breuil–Mézard conjecture, when $p = 2$. We place no restriction on $p$ in this section.

Let $\rho : G_{\mathbb{Q}_{p}} \to \text{GL}_{2}(k)$ be a continuous representation which is either absolutely irreducible, in which case we let $\pi$ be a supersingular representation of $G$ such that $V(\pi) \cong \rho$, or which is isomorphic to $\left( \begin{array}{cc} \chi_{1} & * \\ 0 & \chi_{2} \end{array} \right)$, a nonsplit extension with $\chi_{1}\chi_{2}^{-1} \neq 1$, $\omega^{\pm 1}$, in which case we let $\pi = (\text{Ind}_{B}^{G} \chi_{1} \otimes \chi_{2} \omega^{-1})_{\text{sm}}$. As before we let $R^{\psi}$ be the quotient of the universal deformation ring of $\rho$ parameterizing deformations with determinant $\psi \varepsilon$ and let $\rho^{\text{univ}}$ be the tautological deformation of $\rho$ to $R^{\psi}$.

**Proposition 2.27.** $P$ satisfies the hypotheses (N0)–(N2) of [Paškūnas 2015b, §4].

**Proof.** (N0) says that $k \hat{\otimes}_{R^{\psi}} P$ is of finite length and finitely generated over $O[[K]]$. This follows from Proposition 2.8. To verify (N1) we need to show that

$$\text{Hom}_{\text{SL}_{2}(\mathbb{Q}_{p})}(1, P^{\vee}) = 0.$$
The $\text{SL}_2(\mathbb{Q}_p)$-invariants in $P^\vee$ are stable under the action of $G$. Since $P^\vee$ is an injective envelope of $\pi$, if the subspace is nonzero then it must intersect $\pi$ nontrivially. However, $\pi^{\text{SL}_2(\mathbb{Q}_p)} = 0$, which concludes the proof. (N2) requires $\check{V}(P)$ and $\rho_{\text{univ}}$ to be isomorphic as $R^\psi \llbracket G_{\mathbb{Q}_p} \rrbracket$-modules and this is proved in Theorem 2.22. □

Recall from [Serre 2000, §V.A] that the group of $d$-dimensional cycles $Z_d(A)$ of a noetherian ring $A$ is a free abelian group generated by $p \in \text{Spec } A$ with $\dim A/p = d$. For $d$-dimensional cycles $\sum_p n_pp$ and $\sum_p m_pp$, we write $\sum_p n_pp \leq \sum_p m_pp$, if $n_p \leq m_p$ for all $p \in \text{Spec } A$ with $\dim A/p = d$.

If $M$ is a finitely generated $A$-module of dimension at most $d$ then $M_p$ is an $A_p$-module of finite length, which we denote by $\ell_{A_p}(M_p)$, for all $p$ with $\dim A/p = d$. We note that $\ell_{A_p}(M_p)$ is nonzero only for finitely many $p$. Thus $z_d(M) := \sum_p \ell_{A_p}(M_p)p$, where the sum is taken over all $p \in \text{Spec } A$ such that $\dim A/p = d$, is a well defined element of $Z_d(A)$.

If $(A, m)$ is a local ring then we define a Hilbert–Samuel multiplicity $e(z)$ of a cycle $z = \sum_p n_pp \in Z_d(A)$ to equal $\sum_p n_p e(A/p)$, where $e(A/p)$ is the Hilbert–Samuel multiplicity of the ring $A/p$. If $M$ is a finitely generated $A$-module of dimension $d$ then the Hilbert–Samuel multiplicity of $M$ is equal to the Hilbert–Samuel multiplicity of its cycle $z_d(M)$; see [Serre 2000, §V.2].

If $\Theta$ is a continuous representation of $K$ on a free $\mathcal{O}$-module of finite rank, we let

$$M(\Theta) := \left(\text{Hom}^{\text{cont}}_{\mathcal{O}[K]}(P, \mathcal{O})^{d}\right)^d,$$

where $(*)^d := \text{Hom}_{\mathcal{O}}(\ast, \mathcal{O})$. If $\lambda$ is a smooth representation of $K$ on an $\mathcal{O}$-torsion module of finite length then we let

$$M(\lambda) := \left(\text{Hom}^{\text{cont}}_{\mathcal{O}[K]}(P, \lambda^{\vee})\right)^\vee,$$

where the superscript $\vee$ denotes the Pontryagin dual.

**Proposition 2.28.** Let $\Theta$ be a continuous representation of $K$ on a free $\mathcal{O}$-module of finite rank with central character $\psi$. Then $M(\Theta)$ is a finitely generated $R^\psi$-module. If $M(\Theta)$ is nonzero then it is Cohen–Macaulay and has Krull dimension equal to 2. We have an equality of 1-dimensional cycles

$$z_1(M(\Theta)/\sigma) = \sum_{\sigma} m_\sigma z_1(M(\sigma)), \quad (27)$$

where the sum is taken over all the irreducible smooth $k$-representations of $K$, and $m_\sigma$ denotes the multiplicity with which $\sigma$ appears as a subquotient of $\Theta \otimes_{\mathcal{O}} k$.

Moreover, $M(\sigma) \neq 0$ if and only if $\text{Hom}_K(\sigma, \pi) \neq 0$, in which case the Hilbert–Samuel multiplicity of $z_1(M(\sigma))$ is equal to 1.

**Proof.** We showed in Proposition 2.27 that $k \otimes_{R^\psi} P$ is a finitely generated $\mathcal{O}[K]$-module. It follows from Corollary 2.5 in [Paškūnas 2015b] that $M(\Theta)$ is a finitely
 generated \( R^\psi \)-module. The restriction of \( P \) to \( K \) is projective in \( \text{Mod}^\text{pro}_{K, \psi}(O) \) by [Paškūnas 2015b, Corollary 5.3]. Proposition 2.24 in [Paškūnas 2015b] implies that (27) holds as an equality of \((d - 1)\)-dimensional cycles, where \( d \) is the Krull dimension of \( M(\Theta) \). Theorem 5.2 in [Paškūnas 2015b] shows that there is an \( x \) in the maximal ideal of \( R^\psi \) such that we have an exact sequence \( 0 \to P \to P/xP \to 0 \), where the restriction of \( P/xP \) to \( K \) is a projective envelope of \((\text{soc}_K \pi)^\vee \) in \( \text{Mod}^\text{pro}_{K, \psi}(O) \). Lemma 2.33 in [Paškūnas 2015b] implies that \( M(\Theta) \) is a Cohen–Macaulay module of dimension 2 and that \( \sigma, x \) is a regular sequence of parameters. If \( \sigma \) is an irreducible smooth \( k \)-representation of \( K \) with central character \( \psi \) then the proof of [Paškūnas 2015b, Lemma 2.33] yields an exact sequence

\[
0 \to M(\sigma) \to M(\sigma) \to \left( \text{Hom}_{O[[K]]}(P/xP, \sigma^\vee) \right)^\vee \to 0.
\]

Since \( P/xP \) is a projective envelope of \((\text{soc}_K \pi)^\vee \) in \( \text{Mod}^\text{pro}_{K, \psi}(O) \), we deduce that \( \text{dim}_k M(\sigma)/xM(\sigma) \) is equal to \( \dim_k \text{Hom}_K(\sigma, \pi) \). If \( \text{Hom}_K(\sigma, \pi) \) is zero then Nakayama’s lemma implies that \( M(\sigma) = 0 \). If \( \text{Hom}_K(\sigma, \pi) \) is nonzero then it is a one-dimensional \( k \)-vector space, since the \( K \)-socle of \( \pi \) is multiplicity free. The exact sequence \( 0 \to M(\sigma) \to M(\sigma) \to k \to 0 \) implies that \( M(\sigma) \) is a cyclic module, and if \( a \) denotes its annihilator then \( R^\psi/a \cong k[[x]] \).

\[ \square \]

**Remark 2.29.** If \( \rho \) is absolutely irreducible and \( \rho|_{I_{Q_p}} \cong (\omega_2^{r+1} \oplus \omega_2^{p(r+1)}) \otimes \omega^m \) then

\[
\text{soc}_K \pi \cong \left( \text{Sym}^r k^2 \oplus \text{Sym}^{p-1-r} k^2 \otimes \det^r \right) \otimes \det^m,
\]

where \( 0 \leq r \leq p - 1, \ 0 \leq m \leq p - 2 \) and \( \omega_2 \) is the fundamental character of Serre of niveau 2; see [Breuil 2003a; 2003b]. If \( \rho \cong \left( \begin{smallmatrix} \chi_1 & \ast \\ 0 & \chi_2 \omega^{r+1} \end{smallmatrix} \right) \otimes \omega^m \), where \( \chi_1, \chi_2 \) are unramified and \( \chi_1 \neq \chi_2 \omega^{r+1} \) then

\[
\pi \cong \left( \text{Ind}_B^G \chi_1 \otimes \chi_2 \omega^r \right)_{\text{sm}} \otimes \omega^m \circ \det.
\]

Hence, \( \text{soc}_K \pi \cong \text{Sym}^r k^2 \otimes \det^m \) if \( 0 < r < p - 1 \) and \( \det^m \oplus \text{Sym}^{p-1} k^2 \otimes \det^m \) otherwise. In particular, \( \text{soc}_K \pi \) is multiplicity free.

If \( n \in m-\text{Spec } R^\psi[1/p] \) then the residue field \( \kappa(n) \) is a finite extension of \( L \). Let \( \mathcal{O}_{\kappa(n)} \) be the ring of integers in \( \kappa(n) \). By specializing the universal deformation at \( n \), we obtain a continuous representation \( \rho_n^{\text{univ}} : G_{\mathbb{Q}_p} \to \text{GL}_2(\mathcal{O}_{\kappa(n)}) \), which reduces to \( \rho \) modulo the maximal ideal of \( \mathcal{O}_{\kappa(n)} \). A \( p \)-adic Hodge type \((w, \tau, \psi)\) consists of the following data: \( w = (a, b) \) is a pair of integers with \( b > a \), \( \tau : I_{Q_p} \to \text{GL}_2(L) \) is a representation of the inertia subgroup with an open kernel and \( \psi : G_{\mathbb{Q}_p} \to \mathcal{O}_x^\times \) is a continuous character such that \( \psi \varepsilon \equiv \det \rho \mod \sigma \), \( \psi|_{I_{Q_p}} = \varepsilon^{a+b-1} \det \tau \), where \( \varepsilon \) is the \( p \)-adic cyclotomic character. If \( \rho_n^{\text{univ}} \) is potentially semistable then we say that it is of type \((w, \tau, \psi)\) if its Hodge–Tate weights are equal to \( w \), the determinant
is equal to $\psi$ and the restriction of the Weil–Deligne representation, associated to $\rho_n^{\text{univ}}$ by Fontaine [1994], to $I_{\Omega_p}$ is isomorphic to $\tau$.

Henniart [2002] has shown the existence of a smooth irreducible representation $\sigma(\tau)$ (resp. $\sigma^{\text{cr}}(\tau)$) of $K$ on an $L$-vector space such that if $\pi$ is a smooth absolutely irreducible infinite-dimensional representation of $G$ and $\text{LL}(\pi)$ is the Weil–Deligne representation attached to $\pi$ by the classical local Langlands correspondence then $\text{Hom}_K(\sigma(\tau), \pi) \neq 0$ (resp. $\text{Hom}_K(\sigma^{\text{cr}}(\tau), \pi) \neq 0$) if and only if $\text{LL}(\pi)|_{I_{\Omega_p}} \cong \tau$ (resp. $\text{LL}(\pi)|_{I_{\Omega_p}} \cong \tau$ and the monodromy operator $N$ is 0). The representations $\sigma(\tau)$ and $\sigma^{\text{cr}}(\tau)$ are uniquely determined if $p > 2$. If $p = 2$ there might be different choices; we choose one.

We let $\sigma(\omega, \tau) := \sigma(\tau) \otimes \text{Sym}^{b-a-1} L^2 \otimes \det^a$. Then $\sigma(\omega, \tau)$ is a finite-dimensional $L$-vector space. Since $K$ is compact and the action of $K$ on $\sigma(\omega, \tau)$ is continuous, there is a $K$-invariant $\mathcal{O}$-lattice $\Theta$ in $\sigma(\omega, \tau)$. Then $\Theta/(\omega)$ is a smooth finite-length $k$-representation of $K$, and we let $\overline{\sigma(\omega, \tau)}$ be its semisimplification. One may show that $\overline{\sigma(\omega, \tau)}$ does not depend on the choice of a lattice. For each smooth irreducible $k$-representation $\sigma$ of $K$ we let $m_\sigma(\omega, \tau)$ be the multiplicity with which $\sigma$ occurs in $\overline{\sigma(\omega, \tau)}$. We let $\sigma^{\text{cr}}(\omega, \tau) := \sigma^{\text{cr}}(\tau) \otimes \text{Sym}^{b-a-1} L^2 \otimes \det^a$ and let $m^{\text{cr}}_\sigma(\omega, \tau)$ be the multiplicity of $\sigma$ in $\overline{\sigma^{\text{cr}}(\omega, \tau)}$. If $p = 2$ then one may show that $\overline{\sigma(\omega, \tau)}$ and $\overline{\sigma^{\text{cr}}(\omega, \tau)}$ do not depend on the choice of $\sigma(\tau)$ and $\sigma^{\text{cr}}(\tau)$.

**Proposition 2.30.** Let $V = \sigma(\omega, \tau)$ (resp. $V = \sigma^{\text{cr}}(\omega, \tau)$) and let $\Theta$ be a $K$-invariant lattice in $V$. Then $n \in \text{m-Spec } R^\psi[1/p]$ lies in the support of $M(\Theta)$ if and only if $\rho_n^{\text{univ}}$ is potentially semistable (resp. potentially crystalline) of type $(\omega, \tau, \psi)$. Moreover, for such $n$, we have $\dim_{\kappa(n)} M(\Theta) \otimes_{R^\psi} \kappa(n) = 1$.

**Proof.** Proposition 2.22 of [Paškūnas 2015b] implies that

$$\dim_{\kappa(n)} M(\Theta) \otimes_{R^\psi} \kappa(n) = \dim_{\kappa(n)} \text{Hom}_K(V, \Pi(\kappa(n))).$$

Since $V$ is a locally algebraic representation,

$$\text{Hom}_K(V, \Pi(\kappa(n))) \cong \text{Hom}_K(V, \Pi(\kappa(n))^{\text{alg}}),$$

where the superscript alg denotes the subspace of locally algebraic vectors. This last subspace is nonzero if and only if $\rho_n^{\text{univ}}$ is potentially semistable (resp. potentially crystalline) of type $(\omega, \tau, \psi)$, in which case it is one-dimensional. The argument is identical to the proof of [Paškūnas 2015b, Proposition 4.14], except that, because we assume that $\rho$ is generic, we don’t have to consider the nasty cases here. □

**Corollary 2.31.** There exists a reduced, $\mathcal{O}$-torsion-free quotient $R^\psi(\omega, \tau)$ of $R^\psi$ such that a map of $\mathcal{O}$-algebras $x : R^\psi \to L'$ into a finite field extension of $L$ factors through $R^\psi(\omega, \tau)$ if and only if $\rho_x^{\text{univ}}$ is potentially semistable of type $(\omega, \tau, \psi)$.

Moreover, if $\Theta$ is a $K$-invariant $\mathcal{O}$-lattice in $\sigma(\omega, \tau)$ and $a$ is the $R^\psi$-annihilator of $M(\Theta)$ then $R^\psi(\omega, \tau) = R^\psi/\sqrt{a}$. 
The same result holds if we consider potentially crystalline instead of potentially semistable representations with $\sigma^{cr}(w, \tau)$ instead of $\sigma(w, \tau)$.

Proof. Since the support of $M(\Theta)$ is closed in $\text{Spec } R^\psi$, the assertion follows from Proposition 2.30. □

Corollary 2.32. Let $\Theta$ be a $K$-invariant lattice in either $\sigma(w, \tau)$ or $\sigma^{cr}(w, \tau)$ and let $a$ be the $R^\psi$-annihilator of $M(\Theta)$. Then we have equalities of cycles

$$z_2(R^\psi/a) = z_2(M(\Theta)), \quad z_1(R^\psi/(a, \sigma)) = z_1(M(\Theta)/\sigma).$$

Proof. The last part of Proposition 2.30 implies that $M(\Theta)$ is generically free of rank 1. This implies the first assertion; see [Paškūnas 2015b, Lemma 2.27]. The second follows from the first combined with the fact that $\sigma$ is both $R^\psi/a$- and $M(\Theta)$-regular; see Proposition 2.2.13 in [Emerton and Gee 2014]. □

Proposition 2.33. Let $a$ be the $R^\psi$-annihilator of $M(\Theta)$, where $\Theta$ is a $K$-invariant $\mathcal{O}$-lattice in $\sigma(w, \tau)$ (resp. $\sigma^{cr}(w, \tau)$). Then $R^\psi/a$ is reduced. In particular, it is equal to $R^\psi(w, \tau)$ (resp. $R^{\psi, cr}(w, \tau)$).

Proof. Proposition 2.30 of [Paškūnas 2015b] together with the last part of Proposition 2.30 of the current paper says that it is enough to show that, for almost all $n$ in $m$-$\text{Spec } R^\psi[1/p]$ lying in the support of $M(\Theta)$,

$$\dim_{\kappa(n)} \text{Hom}_K(V, \Pi(R^\psi_n/n^2R^\psi_n)) \leq 2.$$ 

This amounts to checking that the subspace $E$ of $\text{Ext}^1_G(\Pi(\kappa(n)), \Pi(\kappa(n)))$ generated by the extensions of admissible unitary $\kappa(n)$-Banach spaces $0 \to \Pi(\kappa(n)) \to B \to \Pi(\kappa(n)) \to 0$ such that the induced map between the subspaces of locally algebraic vectors $B^{alg} \to \Pi(\kappa(n))^{alg}$ is surjective, is at most one-dimensional; see the proof of [Paškūnas 2015b, Corollary 4.21].

If $\tau$ does not extend to an irreducible representation of $W_{\Omega_p}$ then the proof of [Paškūnas 2015b, Theorem 4.19] carries over: the key input into that proof is that the closure of $\Pi(\kappa(n))^{alg}$ in $\Pi(\kappa(n))$ is equal to the universal unitary completion of $\Pi(\kappa(n))^{alg}$ and the only case of this fact not covered by the references given in the proof of [Paškūnas 2015b, Theorem 4.19] is when $p = 2$ and $\Pi(\kappa(n))^{alg} \cong (\text{Ind}_B^G \chi \otimes \chi | \cdot |^{-1})_{sm} \otimes W$, where $W$ is an algebraic representation of $G$ and $\chi : \Omega_p^\times \to \kappa(n)^\times$ is a smooth character. However, in that case it is explained in the second paragraph of the proof of [Paškūnas 2014, Proposition 6.13] how to deduce from [Paškūnas 2009, Proposition 4.2] that any $G$-invariant $\mathcal{O}$-lattice in $\Pi(\kappa(n))^{alg}$ is a finitely generated $\mathcal{O}[G]$-module, which provides the key input also in this case. We note that the assumption $p > 2$ in [Paškūnas 2009, §4] is only used to apply the results of Berger, Li and Zhu; in particular, the proof of [Paškūnas 2009, Proposition 4.2] works for all $p$. 

If $\tau$ extends to an irreducible representation of $W_{\Omega_p}$ then the assertion is proved by Dospinescu [2015]. Although the main theorem of [Dospinescu 2015] is stated under the assumption $p \geq 5$, the argument only uses that assumption if we let $\Pi = \Pi(\kappa(n))$, in which case $\det \tilde{V}(\Pi) = \psi \varepsilon$ and $\dim L \text{Ext}_G^1(\Pi, \Pi) = 3$. This is given by Corollaries 2.24 and 2.26.

\textbf{Theorem 2.34.} There is a finite set $\{C_\sigma\}_\sigma \subset Z_1(R^\psi/\sigma)$, indexed by the irreducible smooth $k$-representations $\sigma$ of $K$, such that for all $p$-adic Hodge types $(\omega, \tau)$ we have equalities

$$z_1(R^\psi(\omega, \tau)/\sigma) = \sum_{\sigma} m_{\sigma}(\omega, \tau)C_\sigma,$$

$$z_1(R^\psi,cr(\omega, \tau)/\sigma) = \sum_{\sigma} m_{\sigma}^{cr}(\omega, \tau)C_\sigma.$$

The cycle $C_\sigma$ is nonzero if and only if $\text{Hom}_K(\sigma, \pi) \neq 0$, in which case its Hilbert–Samuel multiplicity is equal to 1.

\textbf{Proof.} Let $a$ be the $R^\psi$-annihilator of $M(2)$, where $\Theta$ is a $K$-invariant $O$-lattice in $\sigma(\omega, \tau)$. Corollary 2.31 and Proposition 2.33 imply that

$$z_1(R^\psi(\omega, \tau)/\sigma) = z_1(R^\psi/(\sqrt{a}, \sigma)) = z_1(R^\psi/(a, \sigma)).$$

Corollary 2.32 and Proposition 2.28 imply that

$$z_1(R^\psi/(a, \sigma)) = \sum_{\sigma} m_{\sigma}(\omega, \tau)z_1(M(\sigma)).$$

We let $C_\sigma = z_1(M(\sigma))$. The proof in the potentially crystalline case is the same. \qed

\textbf{Remark 2.35.} One may use a global argument to prove Proposition 2.33, without using the results of [Dospinescu 2015]. However, one needs to assume that the local residual representation can be realized as a restriction to $G_{\Omega_p}$ of a global modular representation.

Let $b$ be the kernel $R^\psi/a \rightarrow R^\psi/\sqrt{a}$. Since $M(\Theta)$ is Cohen–Macaulay, $R^\psi/a$ is equidimensional. Thus if $b$ is nonzero then it is a 2-dimensional $R^\psi$-module, and the cycle $z_1(b/\sigma)$ is nonzero. Since

$$z_1(R^\psi/(a, \sigma)) = z_1(R^\psi/(\sqrt{a}, \sigma)) + z_1(b/\sigma),$$

if $R^\psi/a$ is not reduced then we would conclude that $e(R^\psi/(a, \sigma)) > e(R^\psi(\omega, \tau)/\sigma)$. Since $e(R^\psi/(a, \sigma)) = e(M(\Theta)/\sigma) = \sum_{\sigma} m_{\sigma}(\omega, \tau)e(C_\sigma)$, in this case we would obtain a contradiction to the Breuil–Mézard conjecture.

If the residual representation can be suitably globalized (when $p = 2$ this means that it is of the form $\tilde{\rho}|_{G_{\Omega_p}}$, where $\tilde{\rho}$ satisfies the assumptions made in Section 3B) then a global argument gives an inequality in the opposite direction, thus allowing

\footnote{I thank G. Dospinescu for pointing this out to me.}
us to conclude that $R^\psi/a$ is reduced. If $p > 2$ then such an argument is made in [Kisin 2009a, §2.3]. If $p = 2$ then the same argument can be made using inequality (41) in the proof of Proposition 3.17 and the proof of Corollary 3.27.

**Remark 2.36.** If $R^\square$ is the framed deformation ring of $\rho$ and $R$ is the universal deformation ring of $\rho$ then $R^\square \cong R[[x_1, x_2, x_3]]$. Thus we have a map of cycle groups

$$f : \mathcal{Z}_i(R) \rightarrow \mathcal{Z}_{i+3}(R^\square), \quad p \mapsto p[[x_1, x_2, x_3]],$$

which preserves Hilbert–Samuel multiplicities. The extra variables only keep track of a choice of basis. This implies that if $R^\psi,\square(w, \tau)$ is the quotient of $R^\square$ parametrizing potentially semistable framed deformations of type $(w, \tau, \psi)$ then $R^\psi,\square(w, \tau) \cong R^\psi(w, \tau)[[x_1, x_2, x_3]]$, so that the cycle of $R^\psi,\square(w, \tau)/\sigma\tau$ is the image of the cycle of $R^\psi(w, \tau)/\sigma\tau$ under $f$. Using this, one may deduce a version of Theorem 2.34 for framed deformation rings.

Let $\rho = (\chi_1^0 \chi_2^0)$, and let $R^\square$ be the universal framed deformation ring of $\rho$. Let $R^\psi,\square(w, \tau)$ (resp. $R^\psi,\square,\text{cr}(w, \tau)$) be the reduced, $O$-torsion-free quotient of $R^\square$ parametrizing potentially semistable (resp. potentially crystalline) lifts of $p$-adic Hodge type $(w, \tau, \psi)$.

**Theorem 2.37.** There is a subset $\{C_1, \sigma, C_2, \sigma\}_\sigma$ of $\mathcal{Z}_4(R^\psi,\square/\sigma\tau)$ indexed by the irreducible smooth $k$-representations $\sigma$ of $K$ such that for all $p$-adic Hodge types $(w, \tau)$ we have equalities

$$z_4(R^\psi,\square(w, \tau)/\sigma\tau) = \sum_\sigma m_\sigma(w, \tau)(C_1, \sigma + C_2, \sigma),$$

$$z_4(R^\psi,\square,\text{cr}(w, \tau)/\sigma\tau) = \sum_\sigma m_{\text{cr}}(w, \tau)(C_1, \sigma + C_2, \sigma).$$

The cycle $C_1, \sigma$ is nonzero if and only if $\text{Hom}_K(\sigma, (\text{Ind}_B^G \chi_1 \otimes \chi_2 \omega^{-1})_{\text{sm}}) \neq 0$, and $C_2, \sigma$ is nonzero if and only if $\text{Hom}_K(\sigma, (\text{Ind}_B^G \chi_2 \otimes \chi_1 \omega^{-1})_{\text{sm}}) \neq 0$, in which case the Hilbert–Samuel multiplicity is equal to 1.

**Proof.** Given Theorem 2.34, the assertion follows from Theorem 7.3 and Remark 7.4 of [Paškūnas 2015a].

The following corollary will be used in the global part of the paper.

**Corollary 2.38.** Assume that $p = 2$, $\psi$ is unramified and either $\rho$ is absolutely irreducible or $\rho^{ss} = \chi_1 \oplus \chi_2$, with $\chi_1 \neq \chi_2$. If $w = (0, 1)$ and $\tau = 1 \oplus 1$ then

$$R^\psi,\square,\text{cr}(w, \tau) = R^\psi,\square(w, \tau).$$

In other words, every semistable lift of $\rho$ with Hodge–Tate weights $(0, 1)$ is crystalline.
Proof. It is enough to prove the statement when \( \rho \) is nonsplit. Since if the assertion was false in the split case then by choosing a different lattice in the semistable, noncrystalline lift we would also obtain a contradiction in the nonsplit case. Since framed deformation rings are formally smooth over the nonframed ones, it is enough to prove that \( R^\psi(w, \tau) = R^\psi,cr(w, \tau) \). By the same argument as in Remark 2.35 we see that it is enough to show that \( R^\psi(w, \tau)/\sigma \) and \( R^\psi,cr(w, \tau)/\sigma \) have the same cycles (and even the equality of Hilbert–Samuel multiplicities will suffice). Since \( p = 2 \) there are only 2 irreducible smooth \( k \)-representations of \( K: 1 \) and \( \text{st} \). The \( K \)-socle of \( \pi \) in all the cases is isomorphic to \( 1 \oplus \text{st} \), \( \sigma(w, \tau)/\sigma \cong \text{st} \) and \( \sigma^{cr}(w, \tau)/\sigma \). The assertion follows from Theorem 2.34. \( \square \)

Remark 2.39. Assume that \( p = 2 \), let \( \xi : G_{Q_p} \to O^\times \) be unramified and congruent to \( \psi \) modulo \( \sigma \), and let \( (w, \tau) \) be arbitrary. It follows from Theorem 2.34, Remark 2.36, Theorem 2.37 and the proof of Corollary 2.38 that

\[
z_4\left(R^\psi,\square(w, \tau)/\sigma\right) = (m_1(w, \tau) + m_{\text{st}}(w, \tau))z_4\left(R^{\xi,\square}((0, 1), 1 \oplus 1)/\sigma\right),
\]

where the cycles live in \( Z_4(R^{\square}) \). This equality implies the equality of the respective Hilbert–Samuel multiplicities.

3. Global part

In the global part of the paper we let \( p = 2 \), so that \( L \) is a finite extension of \( Q_2 \) with the ring integers \( O \) and residue field \( k \).

3A. Quaternionic modular forms. We follow very closely [Kisin 2009b, §3.1]. Let \( F \) be a totally real field in which 2 splits completely. Let \( D \) be a quaternion algebra with center \( F \), ramified at all the infinite places of \( F \) and a set of finite places \( \Sigma \) which does not contain any primes dividing 2. We fix a maximal order \( O_D \) of \( D \), and for each finite place \( v \notin \Sigma \) we have an isomorphism \( (O_D)_v \cong M_2(O_{F_v}) \). For each finite place \( v \) of \( F \) we will denote by \( N(v) \) the order of the residue field at \( v \), and by \( \sigma_v \in F_v \) a uniformizer.

Denote by \( \mathbb{A}_F^f \subset \mathbb{A}_F \) the finite adeles, and let \( U = \prod_v U_v \) be a compact open subgroup contained in \( \prod_v (O_D)_v^\times \). We assume that if \( v \in \Sigma \) then \( U_v = (O_D)_v^\times \) and if \( v \mid 2 \) then \( U_v = \text{GL}_2(O_{F_v}) = \text{GL}_2(\mathbb{Z}_2) \). Let \( A \) be a topological \( \mathbb{Z}_2 \)-algebra. For each \( v \mid 2 \), we fix a continuous representation \( \sigma_v : U_v \to \text{Aut}(W_{\sigma_v}) \) on a finite free \( A \)-module. Write \( W_{\sigma} = \bigotimes_{v \mid 2} W_{\sigma_v} \) and denote by \( \sigma : \prod_{v \mid 2} U_v \to \text{Aut}(W_{\sigma}) \) the corresponding representation. We regard \( \sigma \) as being a representation of \( U \) by letting \( U_v \) act trivially if \( v \nmid 2 \). Finally, assume there exists a continuous character \( \psi : (\mathbb{A}_F^f)^\times/F^\times \to A^\times \) such that, for any place \( v \) of \( F \), the action of \( U_v \cap O_{F_v}^\times \) on \( \sigma \) is given by multiplication by \( \psi \). We extend the action of \( U \) on \( W_{\sigma} \) to \( U(\mathbb{A}_F^f)^\times \) by letting \( (\mathbb{A}_F^f)^\times \) act via \( \psi \).
Let $S_{\sigma, \psi}(U, A)$ denote the set of continuous functions

$$f : D^\times \setminus (D \otimes_F \mathbb{A}_F^\times)^\times \to W_\sigma$$

such that for $g \in (D \otimes_F \mathbb{A}_F^\times)^\times$ we have $f(gu) = \sigma(u)^{-1}f(g)$, $u \in U$, and $f(gz) = \psi^{-1}(z)f(g)$, $z \in (\mathbb{A}_F^\times)^\times$. If we write $(D \otimes_F \mathbb{A}_F^\times)^\times = \bigoplus_{i \in I} D^\times t_i U (\mathbb{A}_F^\times)^\times$ for some $t_i \in (D \otimes_F \mathbb{A}_F^\times)^\times$ and some finite index set $I$, then we have an isomorphism of $A$-modules

$$S_{\sigma, \psi}(U, A) \overset{\cong}{\to} \bigoplus_{i \in I} W_\sigma(U(\mathbb{A}_F^\times)^\times \cap t_i^{-1} D^\times t_i / F^\times), \quad f \mapsto (f(t_i))_{i \in I}. \quad (28)$$

**Lemma 3.1.** Let $U_{\text{max}} = \prod_v \mathcal{O}_{D_v}^\times$, where the product is taken over all finite places of $F$. Let $t \in (D \otimes_F \mathbb{A}_F^\times)^\times$. Then the group $(U_{\text{max}}(\mathbb{A}_F^\times)^\times \cap t D^\times t^{-1}) / F^\times$ is finite and there is an integer $N$, independent of $t$, such that its order divides $N$.

**Proof.** This is explained in Section 7.2 of [Khare and Wintenberger 2009b]; see also [Taylor 2006, Lemma 1.1]. \[\square\]

I thank Mark Kisin for explaining the proof of the following lemma to me.

**Lemma 3.2.** Let $v_1$ be a finite place of $F$ such that $D$ splits at $v_1$ and $v_1$ does not divide $2N$, where $N$ is the integer defined in Lemma 3.1. Let $U = \prod_v U_v$ be a subgroup of $(D \otimes_F \mathbb{A}_F^\times)^\times$ such that $U_v = \mathcal{O}_{D_v}^\times$ if $v \neq v_1$ and $U_{v_1}$ is the subgroup of upper triangular, unipotent matrices modulo $\mathcal{O}_{v_1}$. Then

$$(U(\mathbb{A}_F^\times)^\times \cap t D^\times t^{-1}) / F^\times = 1 \quad \text{for all } t \in (D \otimes_F \mathbb{A}_F^\times)^\times. \quad (29)$$

**Proof.** Let $u \in (U(\mathbb{A}_F^\times)^\times \cap t D^\times t^{-1})$ such that $u \notin F^\times$. Then the $F$-subalgebra $F[u]$ of $t D^\times t^{-1}$ is a quadratic field extension of $F$. Let $u'$ be the conjugate of $u$ over $F$. Then $u' = \text{Nm}(u)/u$, where $\text{Nm}$ is the reduced norm. Consider $w = u/u' = u^2/\text{Nm}(u)$. Write $u = hg$ with $h \in U$ and $g \in (\mathbb{A}_F^\times)^\times$. Then $\text{Nm}(g) = g^2$ and so $w = u/u' = h^2/\text{Nm}(h)$. Thus $w$ is in $U$ and also in $t D^\times t^{-1}$.

Since $(U(\mathbb{A}_F^\times)^\times \cap t D^\times t^{-1}) / F^\times$ is a subgroup of $(U_{\text{max}}(\mathbb{A}_F^\times)^\times \cap t D^\times t^{-1}) / F^\times$, $u^N$ is in $F^\times$ and hence $w^N = u^N/(u')^N = 1$. Let $l$ be the prime dividing $N(v_1)$. Since $U_{v_1}$ is a pro-$l$ group and $l$ does not divide $N$, the image of $w$ under the projection $U \to U_{v_1}$ is equal to 1. Since for every $v$ the map $D \to D_v$ is injective, we conclude that $w = 1$, which implies that $u \in F$. \[\square\]

If (29) holds then it follows from (28) that $\sigma \mapsto S_{\sigma, \psi}(U, A)$ defines an exact functor from the category of continuous representations of $U$ on finitely generated $A$-modules, on which $U_v$ for $v \mid 2$ acts trivially and $U \cap (\mathbb{A}_F^\times)^\times$ acts by $\psi$, to the category of finitely generated $A$-modules.

Let $S$ be a finite set of places of $F$ containing $\Sigma$, all the places above 2, all the infinite places and all the places $v$ for which $U_v$ is not maximal. Let $T_{S,A}^{\text{univ}} = A[T_v, S_v]_{v \notin S}$ be a commutative polynomial ring in the indicated formal
variables. We let \((D \otimes F \mathbb{A}^f_F)\times\) act on the space of continuous \(W_\sigma\)-valued functions on \((D \otimes F \mathbb{A}^f_F)\times\) by right translations, \((hf)(g) := f(gh)\). Then \(S_{\sigma, \psi}(U, A)\) becomes a \(\mathbb{T}^{\text{univ}}_{S, A}\)-module with \(S_v\) acting via the double coset \(U_v \left( \begin{smallmatrix} \sigma_v & 0 \\ 0 & 1 \end{smallmatrix} \right) U_v\) and \(T_v\) acting via the double coset \(U_v \left( \begin{smallmatrix} \sigma_v & 0 \\ 0 & 0 \end{smallmatrix} \right) U_v\). We write \(\mathbb{T}_{\sigma, \psi}(U, A)\) or \(\mathbb{T}_{\sigma, \psi}(U)\) for the image of \(\mathbb{T}^{\text{univ}}_{S, A}\) in the endomorphisms of \(S_{\sigma, \psi}(U, A)\).

3B. Residual Galois representation. Keeping the notation of the previous section we fix an algebraic closure \(\overline{F}\) of \(F\) and let \(G_{F, S}\) be the Galois group of the maximal extension of \(F\) in \(\overline{F}\) which is unramified outside \(S\). We view \(\psi\) as a character of \(G_{F, S}\) via global class field theory, normalized so that uniformizers are mapped to geometric Frobenii. Let \(\chi_{\text{cyc}} : G_{F, S} \to \mathcal{O}^\times\) be the global 2-adic cyclotomic character. We note that \(\chi_{\text{cyc}}\) is trivial modulo \(\varpi\). For each place \(v\) of \(F\), including the infinite places, we fix an embedding \(F \hookrightarrow \mathbb{F}_v\). This induces a continuous homomorphism of Galois groups \(G_{F, v} := \text{Gal}(\mathbb{F}_v/F_v) \to G_{F, S}\). We fix a continuous representation \(\overline{\rho} : G_{F, S} \to \text{GL}_2(k)\) and assume that the following conditions hold:

- The image of \(\overline{\rho}\) is nonsolvable.
- \(\overline{\rho}\) is unramified at all finite places \(v \nmid 2\).
- If \(v \in S\) is a finite place, \(v \not\in \Sigma\), and \(v \nmid 2\), then the eigenvalues of \(\overline{\rho}(\text{Frob}_v)\) are distinct.
- If \(v \in \Sigma\) then the eigenvalues of \(\overline{\rho}(\text{Frob}_v)\) are equal.
- \(\det \overline{\rho} \equiv \psi \chi_{\text{cyc}} \pmod{\varpi}\).
- If \(v \in S\) is a finite place, \(v \not\in \Sigma\), and \(v \nmid 2\), then

\[
U_v = \left\{ g \in \text{GL}_2(\mathcal{O}_{F_v}) : g \equiv \left( \begin{smallmatrix} 1 & * \\ 0 & 1 \end{smallmatrix} \right) \pmod{\varpi_v} \right\}
\]

and at least one such \(v\) does not divide \(2N\), so that the condition of Lemma 3.2 is satisfied.

3B1. Local deformation rings. We fix a basis of the underlying vector space \(V_k\) of \(\overline{\rho}\). For each \(v \in S\) let \(R^\square_v\) be the framed deformation ring of \(\overline{\rho}|_{G_{F, v}}\) and let \(R^\psi, \square_v\) be the quotient of \(R^\square_v\) parametrizing lifts with determinant \(\psi \chi_{\text{cyc}}\). We will now introduce some quotients of \(R^\psi, \square_v\).

For \(v \mid 2\) let \(\tau_v\) be a 2-dimensional representation of the inertia group \(I_v\) with an open kernel, and let \(w_v = (a_v, b_v)\) be a pair of integers with \(b_v > a_v\). Let \(\sigma(\tau_v)\) be any absolutely irreducible representation of \(U_v = \text{GL}_2(\mathbb{Z}_2)\) with the property that, for all irreducible infinite-dimensional smooth representations \(\pi\) of \(\text{GL}_2(\mathbb{Q}_2)\), \(\text{Hom}_{U_v}(\sigma(\tau_v), \pi) \neq 0\) if and only if the restriction to \(I_v\) of the Weil–Deligne
representation $LL(\pi)$ associated to $\pi$ via the local Langlands correspondence is isomorphic to $\tau$. The existence of such $\sigma(\tau_v)$ is shown in [Henniart 2002], where it is also shown that if $\text{Hom}_{U_v}(\sigma(\tau_v), \pi) \neq 0$ then it is one-dimensional. We choose a $U_v$-invariant $O$-lattice $\sigma(\tau_v)^0$ in $\sigma(\tau_v)$ and let

$$\sigma_v := \sigma(\tau_v)^0 \otimes_O \text{Sym}^{b_v-a_v-1} O^2 \otimes_O \det^{a_v}.$$  

(30)

We let $R_v^{\psi, \square}(\sigma_v)$ be the reduced, $O$-flat quotient of $R_v^{\psi, \square}$ parametrizing potentially semi-stable lifts with Hodge–Tate weights $w_v$ and inertial type $\tau_v$. This ring is denoted by $R_v^{\psi, \square}(w, \tau)$ in the local part of the paper.

We similarly define $\sigma^{cr}(\tau_v)$ by additionally requiring that $\text{Hom}_{U_v}(\sigma^{cr}(\tau_v), \pi) \neq 0$ if and only if the monodromy operator $N$ in $LL(\pi)$ is zero and $LL(\pi)|_{I_v} \cong \tau_v$. In this case we let

$$\sigma_v := \sigma^{cr}(\tau_v)^0 \otimes_O \text{Sym}^{b_v-a_v-1} O^2 \otimes_O \det^{a_v}.$$  

(31)

We let $R_v^{\psi, \square}(\sigma_v)$ be the quotient of $R_v^{\psi, \square}$ parametrizing potentially crystalline lifts with Hodge–Tate weights $w_v$ and inertial type $\tau_v$. This ring is denoted by $R_v^{\psi, \square, \text{cr}}(w, \tau)$ in the local part of the paper.

It follows either from the local part of the paper or from [Kisin 2008], where a more general result is proved, that if $R_v^{\psi, \square}(\sigma_v)$ is nonzero then it is equidimensional of Krull dimension 5. Since the residue field of $\mathbb{Z}_2$ has 2 elements, $\sigma(\tau_v)$ need not be unique (see [Henniart 2002, §§A.2.6, A.2.7]); however, the semisimplification of $\sigma(\tau_v)^0 \otimes_O k$ is the same in all cases.

If $v$ is infinite then $R_v^{\psi, \square}$ is a domain of Krull dimension 3 and $R_v^{\psi, \square}[1/2]$ is regular [Kisin 2009b, Proposition 2.5.6; Khare and Wintenberger 2009b, Proposition 3.1].

If $v$ is finite, $\tilde{\rho}$ is unramified at $v$ and $\tilde{\rho}(\text{Frob}_v)$ has distinct Frobenius eigenvalues, then $R_v^{\psi, \square}$ has Krull dimension 4 and $R_v^{\psi, \square}[1/2]$ is regular. This follows from [Kisin 2009b, Proposition 2.5.4], where it is shown that the dimension is 4 and the irreducible components are regular. Since we assume that the eigenvalues of $\tilde{\rho}(\text{Frob}_v)$ are distinct, $\tilde{\rho}$ cannot have a lift of the form $\gamma \oplus \gamma \chi_{\text{cyc}}$. It follows from the proof of [Kisin 2009b, Proposition 2.5.4] that different irreducible components of $R_v^{\psi, \square}[1/2]$ do not intersect.

If $v$ is finite, $\psi$ and $\tilde{\rho}$ are unramified at $v$ and $\tilde{\rho}(\text{Frob}_v)$ has equal eigenvalues, then for an unramified character $\gamma : G_{F_v} \rightarrow O^\times$ such that $\gamma^2 = \psi|_{G_{F_v}}$ we let $R_v^{\psi, \square}(\gamma)$ be a reduced $O$-torsion-free quotient of $R_v^{\psi, \square}$ with the property that if $L'/L$ is a finite extension then there is a map $x : R_v^{\psi, \square} \rightarrow L'$ factors through $R_v^{\psi, \square}(\gamma)$ if and only if $V_x$ is isomorphic to $(\gamma \chi_{\text{cyc}}^*)$. It follows from [Kisin 2009b, Proposition 2.5.2] via [Kisin 2009c, Proposition 2.6.6] and [Khare and Wintenberger 2009b, Theorem 3.1] that $R_v^{\psi, \square}(\gamma)$ is a domain of Krull dimension 4 and $R_v^{\psi, \square}(\gamma)[1/2]$ is regular. If $L$ is large enough then there are precisely two such characters, which we denote by $\gamma_1$
and $\gamma_2$. We let $\overline{R}_v^{\psi, \square}$ be the image of

$$R_v^{\psi, \square} \to R_v^{\psi, \square}(\gamma_1)[\frac{1}{2}] \times R_v^{\psi, \square}(\gamma_2)[\frac{1}{2}].$$

Then $\overline{R}_v^{\psi, \square}$ is a reduced, $\mathcal{O}$-flat quotient of $R_v^{\psi, \square}$ such that if $L' / L$ is a finite extension then a map $x : R_v^{\psi, \square} \to L'$ factors through $\overline{R}_v^{\psi, \square}$ if and only if $V_x$ is isomorphic to $(\gamma \chi_{\text{cyc}})^* \otimes \mathbb{Q}$ for an unramified character $\gamma$. Moreover,

$$\overline{R}_v^{\psi, \square}[\frac{1}{2}] \cong R_v^{\psi, \square}(\gamma_1)[\frac{1}{2}] \times R_v^{\psi, \square}(\gamma_2)[\frac{1}{2}].$$

Thus $\overline{R}_v^{\psi, \square}[\frac{1}{2}]$ is regular and equidimensional and the Krull dimension of $\overline{R}_v^{\psi, \square}$ is 4. We let

$$R_S^{\square} = \bigotimes_{v \in S} R_v^{\square}, \quad R_S^{\psi, \square} = \bigotimes_{\sigma \in \Sigma} R_v^{\psi, \square}, \quad \sigma := \bigotimes_{v \mid 2} \sigma_v,$$

and

$$R_S^{\psi, \square}(\sigma) := \bigotimes_{v \mid 2} R_v^{\psi, \square}(\sigma_v) \bigotimes_{v \in S} \overline{R}_v^{\psi, \square} \bigotimes_{v \mid \infty} R_v^{\psi, \square}.$$

It follows from above that $R_S^{\psi, \square}(\sigma)$ is equidimensional of Krull dimension equal to

$$1 + 4 \sum_{v \mid 2} 1 + 3 |\Sigma| + 3 \left(|S| - |\Sigma| - \sum_{v \mid 2} 1 - \sum_{v \mid \infty} 1\right) + 2 \sum_{v \mid \infty} 1 = 1 + 3 |S|. \quad (32)$$

**3B2. Global deformation rings.** Since $\hat{\rho}$ is assumed to have nonsolvable image, $\hat{\rho}$ is absolutely irreducible. We define $R_{F, S}$ to be the quotient of the universal deformation ring of $\hat{\rho}$ parametrizing deformations with determinant $\psi \chi_{\text{cyc}}$. If $Q$ is a finite set of places of $F$ disjoint from $S$ then we let $S_Q = S \cup Q$ and define $R_{F, S_Q}$ in the same way by viewing $\hat{\rho}$ as a representation of $G_{F, S_Q}$.

Denote by $R_{F, S_Q}$ the complete local $\mathcal{O}$-algebra representing the functor which assigns to an artinian, augmented $\mathcal{O}$-algebra $A$ the set of isomorphism classes of tuples $\{V_A, \beta_w\}_{w \in S}$, where $V_A$ is a deformation of $\hat{\rho}$ to $A$ with determinant $\psi \chi_{\text{cyc}}$ and $\beta_w$ is a lift of a chosen basis of $V_k$ to a basis of $V_A$. The map $\{V_A, \beta_w\}_{w \in S} \mapsto \{V_A, \beta_v\}$ induces a homomorphism of $\mathcal{O}$-algebras $R_v^{\psi, \square} \to R_{F, S_Q}$ for every $v \in S$ and hence a homomorphism of $\mathcal{O}$-algebras $R_S^{\psi, \square} \to R_{F, S_Q}$.

**3C. Patching.** For each $n \geq 1$ let $Q_n$ be the set of places of $F$ disjoint from $S$, as in [Kisin 2009b, Lemma 3.2.2] via [Khare and Wintenberger 2009b, Proposition 5.10]. We let $Q_0 = \emptyset$, so that $S_{Q_0} = S$ for $n = 0$. Let $U_{Q_n} = \prod_v (U_{Q_n})_v$ be a compact open subgroup of $(D \otimes_F \mathbb{A}_F^\infty)^\times$ such that $(U_{Q_n})_v = U_v$ for $v \notin Q_n$ and $(U_{Q_n})_v$ is defined as in [Kisin 2009b, §3.1.6] for $v \in Q_n$.

Let $m$ be a maximal ideal of $\mathbb{T}_{S, \mathcal{O}}^\text{univ}$ such that the residue field is $k$, $T_v$ is mapped to $\text{tr} \hat{\rho}(\text{Frob}_v)$ and $S_v$ is mapped to the image of $\psi(\text{Frob}_v)$ in $k$ for all $v \notin S$. We define
with the convention that if \( m \) or \( n \) is given by either (30) or (31). We assume that \( S_{\sigma, \psi}(U, \mathcal{O})_m \neq 0 \). Then for all \( n \geq 0 \) there is a surjective homomorphism of \( \mathcal{O} \)-algebras \( R_{F,S,Q_n}^\psi \rightarrow \prod_{v \in S_{F,Q_n}} (U_{Q_n})_{m,n} \) such that for all \( v \notin S_{Q_n} \) the trace of \( \text{Frob}_v \) of the tautological \( R_{F,S,Q_n}^\psi \)-representation of \( G_{F,S,Q_n} \) is mapped to \( T_v \). Set
\[
M_n(\sigma) = R_{F,S,Q_n}^\psi \times R_{F,S,Q_n}^\psi S_{\sigma, \psi}(U_{Q_n}, \mathcal{O})_{m,n},
\]
with the convention that if \( n = 0 \) then \( Q_n = \emptyset \), \( S_{Q_n} = S \), \( m_{Q_n} = m \), so that
\[
M_0(\sigma) = R_{F,S}^\psi \times R_{F,S}^\psi S_{\sigma, \psi}(U, \mathcal{O})_m.
\]
It follows from the local-global compatibility of Jacquet–Langlands and Langlands correspondences that the action of \( R_{F,S,Q_n}^\psi \) on \( M_n(\sigma) \) factors through the quotient
\[
R_{F,S,Q_n}^\psi(\sigma) := R_{F,S}^\psi(\sigma) \otimes R_{F,S}^\psi R_{F,S,Q_n}^\psi.
\]
Let \( h = \dim_k H^1(G_F, \text{ad} \hat{\rho}) = 2 \cdot |Q_n| \). Let \( a_\infty \) denote the ideal of \( \mathcal{O}[[y_1, \ldots, y_h]] \) generated by \( (y_1, \ldots, y_h) \). Since \( R_{F,S,Q_n}^\psi \) is formally smooth over \( R_{F,S,Q_n}^\psi \) of relative dimension \( j = 4|S| - 1 \) we may choose an identification
\[
R_{F,S,Q_n}^\psi = R_{F,S,Q_n}^\psi [[y_{h+1}, \ldots, y_{h+j}]]
\]
and regard \( M_n(\sigma) \) as an \( \mathcal{O}[[y_1, \ldots, y_{h+j}]] \)-module. This allows us to consider \( R_{F,S,Q_n}^\psi \) as an \( R_S^\psi \)-algebra via the map \( R_S^\psi \rightarrow R_{F,S,Q_n}^\psi / (y_{h+1}, \ldots, y_{h+j}) = R_{F,S,Q_n}^\psi \). We let
\[
R_{F,S,Q_n}^\psi(\sigma) := R_S^\psi(\sigma) \otimes R_{F,S}^\psi R_{F,S,Q_n}^\psi.
\]
Let \( g = 2|Q_n| + 1 \) and \( t = 2 - |S| + |Q_n| \) and let \( \hat{\mathbb{G}}_m \) be the completion of the \( \mathcal{O} \)-group \( \mathbb{G}_m \) along the identity section. The patching argument as in [Khare and Wintenberger 2009b, Proposition 9.3] shows that there exist \( \mathcal{O}[[y_1, \ldots, y_{h+j}]] \)-algebras \( R_\infty(\sigma) \) and \( R_\infty(\sigma) \) and an \( R_\infty(\sigma) \)-module \( M_\infty(\sigma) \) with the following properties:

(P1) There are surjections of \( \mathcal{O} \)-algebras
\[
R_S^\psi(\sigma)[[x_1, \ldots, x_g]] \rightarrow R_\infty(\sigma) \rightarrow R_\infty(\sigma).
\]

(P2) There is an isomorphism of \( R_S^\psi(\sigma) \)-algebras
\[
R_\infty(\sigma)/a_\infty R_\infty(\sigma) \xrightarrow{\cong} R_{F,S}^\psi(\sigma)
\]
and an isomorphism of \( R_{F,S}^\psi(\sigma) \)-modules
\[
M_\infty(\sigma)/a_\infty M_\infty(\sigma) \xrightarrow{\cong} M_0(\sigma).
\]

(P3) \( M_\infty(\sigma) \) is finite flat over \( \mathcal{O}[[y_1, \ldots, y_{h+j}]] \).
(P4) Spf $R'_\infty(\sigma)$ is equipped with a free action of $(\hat{\mathbb{G}}_m)^t$, and a $(\hat{\mathbb{G}}_m)^t$-equivariant morphism $\delta : \text{Spf } R'_\infty(\sigma) \to (\hat{\mathbb{G}}_m)^t$, where $(\hat{\mathbb{G}}_m)^t$ acts on itself by the square of the identity map.

(P5) We have $\delta^{-1}(1) = \text{Spf } R_\infty(\sigma) \subset \text{Spf } R'_\infty(\sigma)$, and the induced action of $(\hat{\mathbb{G}}_m[2])^t$ on $\text{Spf } R_\infty(\sigma)$ lifts to $M_\infty(\sigma)$.

If $A$ is a local noetherian ring of dimension $d$ and $M$ is a finitely generated $A$-module, we denote by $e(M, A)$ the coefficient of $x^d$ in the Hilbert–Samuel polynomial of $M$ with respect to the maximal ideal of $A$, multiplied by $d!$. In particular, $e(M, A) = 0$ if $\dim M < \dim A$. If $M = A$ we abbreviate $e(M, A)$ to $e(A)$.

It follows from [Khare and Wintenberger 2009b, Proposition 2.5] that there is a $\hat{\mathbb{G}}_m[2]^t$-equivariant $R_\infty^\text{inv}(\sigma)$ with residue field $k$ such that $\text{Spf } R_\infty^\text{inv}(\sigma) = \text{Spf } R'_\infty(\sigma)/(\hat{\mathbb{G}}_m)^t$. Moreover,

$$R'_\infty(\sigma) = R_\infty^\text{inv}(\sigma) \hat{\otimes}_O \mathcal{O}[Z_2] \cong R_\infty^\text{inv}(\sigma)[[z_1, \ldots, z_t]].$$

This implies that

$$\dim R'_\infty(\sigma) = \dim R_\infty^\text{inv}(\sigma) + t, \quad e(R'_\infty(\sigma)/\sigma) = e(R_\infty^\text{inv}(\sigma)/\sigma).$$

**Lemma 3.3.** There are $a_1, \ldots, a_t \in m_\sigma^\text{inv}$ such that

$$R_\infty(\sigma) \cong \frac{R_\infty^\text{inv}(\sigma)[[z_1]]}{(1 + z_1)^2 - (1 + a_1)} \hat{\otimes} R_\infty^\text{inv}(\sigma) \cdots \hat{\otimes} R_\infty^\text{inv}(\sigma) \frac{R_\infty^\text{inv}(\sigma)[[z_t]]}{(1 + z_t)^2 - (1 + a_t)}. \quad (35)$$

In particular, $R_\infty(\sigma)$ is a free $R_\infty^\text{inv}(\sigma)$-module of rank $2^t$.

**Proof.** It follows from [Khare and Wintenberger 2009b, Lemma 9.4] that $\text{Spf } R_\infty(\sigma)$ is a $(\hat{\mathbb{G}}_m[2])^t$-torsor over $\text{Spf } R_\infty^\text{inv}(\sigma)$. The assertion follows from [SGA 3 II 1970, Exposé VIII, Proposition 4.1].

**Lemma 3.4.** Let $p \in \text{Spec } R_\infty^\text{inv}(\sigma)$. The group $(\hat{\mathbb{G}}_m[2])^t(\mathcal{O})$ acts transitively on the set of prime ideals of $R_\infty(\sigma)$ lying above $p$.

**Proof.** Let us write $X$ for $\text{Spf } R_\infty(\sigma)$ and $G$ for $(\hat{\mathbb{G}}_m[2])^t$. The action of $G$ on $X$ induces an action of $(\pm 1)^t = G(\mathcal{O}) \hookrightarrow G(R_\infty(\sigma))$ on $X(R_\infty(\sigma))$. If $g \in G(\mathcal{O})$ we let $\phi_g \in X(R_\infty(\sigma))$ be the image of $(g, \text{id}_{R_\infty(\sigma)})$. The map $g \mapsto \phi_g$ induces a homomorphism of groups $G(\mathcal{O}) \to \text{Aut}(R_\infty(\sigma))$. Explicitly, if $g = (\epsilon_1, \ldots, \epsilon_t)$, where $\epsilon_i$ is either 1 or $-1$, then $\phi_g$ is $R_\infty^\text{inv}(\sigma)$-linear and maps $1 + z_i$ to $\epsilon_i(1 + z_i)$ for $1 \leq i \leq t$. It follows from (35) that $G(\mathcal{O})$ acts transitively on the set of maximal ideals of $\kappa(p) \otimes R_\infty^\text{inv}(\sigma) R_\infty(\sigma)$.

**Lemma 3.5.** The support of $M_\infty(\sigma)$ in $\text{Spec } R_\infty(\sigma)$ is a union of irreducible components. The Krull dimension of $\text{Spec } R_\infty(\sigma)$ is equal to $h + j + 1$. 

Proof. It follows from part (P3) above that the support of $M_\infty(\sigma)$ is equidimensional of dimension $h + j + 1$. To prove the assertion it is enough to show that the dimension of $R_\infty(\sigma)$ is less than or equal to $h + j + 1$. Using Lemma 3.3, (34), (P1) and (32) we deduce that $\dim R_\infty(\sigma) \leq \dim R^\psi,\square_S(\sigma) + g - t = 3|S| + 1 + g - t = h + j + 1$. □

**Lemma 3.6.** $e(R'_\infty(\sigma)/\sigma) \leq e(R^\psi,\square_S(\sigma)/\sigma)$.

Proof. It follows from (33) and Lemmas 3.3 and 3.5 that

$$\dim R'_\infty(\sigma) = \dim R_\infty(\sigma) + t = t + h + j + 1 = 3|S| + 1 + g,$$

which is also the dimension of $R^\psi,\square_S(\sigma)[x_1, \ldots, x_g]$ by (32). The surjection in (P1) above implies that

$$e(R'_\infty(\sigma)/\sigma) \leq e(R^\psi,\square_S(\sigma)[x_1, \ldots, x_g]/\sigma) = e(R^\psi,\square_S(\sigma)/\sigma).$$

□

**Lemma 3.7.** If $S_{\sigma,\psi}(U, \mathcal{O})_m$ is supported on a closed point $n \in \text{Spec } R^\psi,\square_S(\sigma)[\frac{1}{2}]$ then the localization $R^\psi,\square_S(\sigma)_n$ is a regular ring.

Proof. Since the rings $R^\psi[\frac{1}{2}]$ are regular for all $v | 2$ it is enough to show that $n$ defines a regular point in $\text{Spec } R^\psi,\square_S(\sigma)$ for all $v | 2$. This follows from the proof of Lemma B.5.1 in [Gee and Kisin 2014]. The argument is as follows: if the point is not regular, then it must lie on the intersection of two irreducible components of $\text{Spec } R^\psi,\square_S(\sigma)$, but this would violate the weight–monodromy conjecture for $\text{WD}(\rho_n|_{G_{\mathbb{F}_v}})$; see [Gee and Kisin 2014] for details. □

**Lemma 3.8.** If $S_{\sigma,\psi}(U, \mathcal{O})_m$ is supported on a closed point $n \in \text{Spec } R_\infty(\sigma)[\frac{1}{2}]$ then the localization $R_\infty(\sigma)_n$ is a regular ring.

Proof. Let $n_S$ be the image of $n$ in $\text{Spec } R^\psi,\square_S[x_1, \ldots, x_g]$, let $n'$ be the image of $n$ in $\text{Spec } R'_\infty(\sigma)$ via the maps in (P1), and let $n'^{inv}$ be the image of $n$ in $\text{Spec } R^{inv}_\infty(\sigma)$ via (35). It follows from Lemma 3.7 that $R^\psi,\square_S(\sigma)[x_1, \ldots, x_g]_{n_S}$ is a regular ring. If the map

$$R^\psi,\square_S(\sigma)[x_1, \ldots, x_g]_{n_S} \rightarrow R'_\infty(\sigma)_{n'},$$

(36)

is an isomorphism, then $R'_\infty(\sigma)_{n'}$ is a regular ring. We may assume that $L$ is sufficiently large, so that using (33) we may write $n' = (n'^{inv}, z_1 - a_1, \ldots, z_t - a_t)$ with $a_i \in \sigma \mathcal{O}$ for $1 \leq i \leq t$. The images of $z_1 - a_1, \ldots, z_t - a_t$ in $n'/n^{2}$ are linearly independent. Since

$$R^{inv}_\infty(\sigma)_{n'^{inv}} \cong R'_\infty(\sigma)_{n'}/(z_1 - a_1, \ldots, z_t - a_t)R'_\infty(\sigma)_{n'},$$

we deduce that $R^{inv}_\infty(\sigma)_{n'^{inv}}$ is regular. It follows from (35) that the map

$$R^{inv}_\infty(\sigma)[\frac{1}{2}] \rightarrow R_\infty(\sigma)[\frac{1}{2}]$$

is étale. Hence $R_\infty(\sigma)_n$ is a regular ring.
Auslander–Buchsbaum theorem shows that $M_{(P^3)}$ that $M_{(P^3)} R$. The action of $L_q \sigma, \psi$ localizing further at $n$ ideal $R$ and $S$ to show that dim $\dim_{(\sigma)} M_{(\sigma)} = \dim M_{(\sigma)} - 1$. This leads to a contradiction, as $M_{(\sigma)}$ is a finitely generated $R_{(\sigma)}$-module. □

Lemma 3.9. Let $A$ be a local noetherian ring and let $(x_1, \ldots, x_d)$ be a system of parameters of $A$. If $A$ is equidimensional then every irreducible component of $A$ contains a closed point of $(A/(x_2, \ldots, x_d))[1/x_1]$.

Proof. Let $p$ be an irreducible component of $A$. If $A/(p, x_2, \ldots, x_d)[1/x_1]$ is zero then $x_1$ is nilpotent in $A/(p, x_2, \ldots, x_d)$. Since $(x_1, \ldots, x_d)$ is a system of parameters of $A$, we conclude that $A/(p, x_2, \ldots, x_d)$ is zero dimensional, which implies that dim $A/p \leq d - 1$, contradicting equidimensionality of $A$. □

Lemma 3.10. There is an integer $r$, independent of $\sigma$ and the choices made in the patching process, such that for all $p \in \text{Spec } R_{(\sigma)}$ in the support of $M_{(\sigma)}$ we have

$$\dim_{\kappa(p)} M_{(\sigma)} \otimes_{R_{(\sigma)}} \kappa(p) \geq r,$$

with equality if $p$ is a minimal prime of $R_{(\sigma)}$ in the support of $M_{(\sigma)}$.

Proof. Let $q$ be a minimal prime of $R_{(\sigma)}$ in the support of $M_{(\sigma)}$. It is enough to show that $\dim_{\kappa(q)} M_{(\sigma)} \otimes_{R_{(\sigma)}} \kappa(q)$ is independent of $q$ and $\sigma$. Since

$$M_{(\sigma)}/(y_1, \ldots, y_{h+j}) M_{(\sigma)} \cong S_{\sigma, \psi}(U, O)_m$$

and $S_{\sigma, \psi}(U, O)_m$ is a finitely generated $O$-module, $y_1, \ldots, y_{h+j}, \sigma$ is a system of parameters for $R_{(\sigma)}/q$ and it follows from Lemma 3.9 that there is a maximal ideal $n$ of $R_{(\sigma)}[\frac{1}{2}]$, contained in $V(q)$, such that $S_{\sigma, \psi}(U, O)_n \neq 0$. It follows from (P3) that $M_{(\sigma)}$ is a Cohen–Macaulay module. The same holds for the localization at $n$. Since $R_{(\sigma)}$ is a regular ring by Lemma 3.8, a standard argument with the Auslander–Buchsbaum theorem shows that $M_{(\sigma)}$ is a free $R_{(\sigma)}$-module. By localizing further at $q$ we deduce that

$$\dim_{\kappa(q)} M_{(\sigma)} \otimes_{R_{(\sigma)}} \kappa(q) = \dim_{\kappa(n)} M_{(\sigma)} \otimes_{R_{(\sigma)}} \kappa(n)$$

$$= \dim_{\kappa(n)} S_{\sigma, \psi}(U, O)_m \otimes_{R_{(\sigma)}} \kappa(n).$$

(37)

So it is enough show that $\dim_{\kappa(n)} S_{\sigma, \psi}(U, O)_m \otimes_{R_{(\sigma)}} \kappa(n)$ is independent of $n$ and $\sigma$. The action of $R_{(\sigma)}$ on $S_{\sigma, \psi}(U, O)_m$ factors through the action of the Hecke algebra $\mathbb{T}_{\sigma, \psi}(U)$, which is reduced. Thus $\mathbb{T}_{\sigma, \psi}(U)[\frac{1}{2}]$ is a product of finite field extensions of $L$ and we have

$$S_{\sigma, \psi}(U, O)_m \otimes_{R_{(\sigma)}} \kappa(n) = S_{\sigma, \psi}(U, O)_n = (S_{\sigma, \psi}(U, O)_m \otimes_{O} L)[n].$$
Let \( \pi = \otimes_v \pi_v \) be the automorphic representation of \( (D \otimes_F \mathbb{A}_F^f)^\times \) corresponding to \( f^D \in (S_{\sigma, \psi}(U, \mathcal{O})_m \otimes_{\mathcal{O}} L)[n] \). We assume that \( L \) is sufficiently large. It follows from the discussion in [Kisin 2009c, §3.1.14], relating \( S_{\sigma, \psi}(U, L) \) to the space of classical automorphic forms on \((D \otimes_F \mathbb{A}_F^f)^\times\), that

\[
\dim_L(S_{\sigma, \psi}(U, \mathcal{O})_m \otimes_{\mathcal{O}} L)[n] = \prod_{v \in \mathcal{S}} \dim_L \pi_v^{U_v} \prod_{v | 2} \dim_L \text{Hom}_{U_v}(\sigma(\tau_v), \pi_v).
\]

We claim that the right-hand side of the above equation is equal to \( 2^{\lvert \mathcal{S} \rvert} \lvert \Sigma \cup \{v \mid 2\infty\} \rvert \). The claim will follow from the local-global compatibility of Langlands and Jacquet–Mao. It follows from Lemma 3.4 and (P5) that all \( \rho_v \) is an associated prime and so \( \text{Hom}_{U_v}(\sigma(\tau_v), \pi_v) = 1 \) if \( v \in \Sigma \) then \( \pi_v \) is an unramified character of \( D_v^\times \), and hence \( \dim_L \pi_v^{U_v} = 1 \). If \( v \notin \Sigma \), \( v | 2\infty \), and \( v \notin \Sigma \) then \( D \) is split at \( v \), \( \bar{\rho}|_{G_{F_v}} \) is unramified and \( \bar{\rho}(\text{Frob}_v) \) has distinct eigenvalues. This implies that \( \rho|_{G_{F_v}} \) is an extension of distinct tamely ramified characters \( \psi_1, \psi_2 \) such that \( \psi_1\psi_2^{-1} \neq \chi_{\text{cyc}}^{-1} \). We deduce that \( \pi_v \) is a tamely ramified principal series. Since \( U_v \) is equal to the subgroup of unipotent upper-triangular matrices modulo \( \mathcal{O}_v \), in this case, we deduce that \( \dim_L \pi_v^{U_v} = 2 \). □

**Lemma 3.11.** There is an integer \( r \), independent of \( \sigma \) and the choices made in the patching process, such that for all minimal primes \( p \) of \( R_{\infty}^{\text{inv}}(\sigma) \) in the support of \( M_{\infty}(\sigma) \) we have

\[
\dim_{\kappa(p)} M_{\infty}(\sigma) \otimes_{R_{\infty}^{\text{inv}}(\sigma)} \kappa(p) = 2^r r.
\]

**Proof.** To ease the notation, let us drop \( \sigma \) from it in this proof. Since \( p \) is minimal, it is an associated prime and so \( M_{\infty} \) will contain \( R_{\infty}^{\text{inv}}/p \) as a submodule. Since \( M_{\infty} \) is \( \mathcal{O} \)-torsion-free, this implies that the quotient field \( \kappa(p) \) has characteristic 0. It follows from (35) that \( R_{\infty} \otimes_{R_{\infty}^{\text{inv}}} \kappa(p) \) is étale over \( \kappa(p) \), and so

\[
R_{\infty} \otimes_{R_{\infty}^{\text{inv}}} \kappa(p) \cong \prod_q \kappa(q),
\]

where the product is taken over all prime ideals \( q \) of \( R_{\infty} \) such that \( q \cap R_{\infty}^{\text{inv}} = p \). From this we get

\[
\dim_{\kappa(p)} M_{\infty} \otimes_{R_{\infty}^{\text{inv}}} \kappa(p) = \sum_q [\kappa(q) : \kappa(p)] \dim_{\kappa(q)} M_{\infty} \otimes_{R_{\infty}} \kappa(q).
\]

It follows from Lemma 3.4 and (P5) that all \( q \) appearing in the sum lie in the support of \( M_{\infty} \). Lemma 3.10 implies that \( \dim_{\kappa(q)} M_{\infty} \otimes_{R_{\infty}} \kappa(q) = r \). Thus

\[
\dim_{\kappa(p)} M_{\infty} \otimes_{R_{\infty}^{\text{inv}}} \kappa(p) = r \dim_{\kappa(p)} R_{\infty} \otimes_{R_{\infty}^{\text{inv}}} \kappa(p) = r 2^r,
\]

where the last equality follows from Lemma 3.3. □
Lemma 3.12. Let $A$ be a local noetherian ring, let $M$, $N$ be finitely generated $A$-modules of dimension $d$, and let $x \in A$ be $M$-regular and $N$-regular. If $\ell_{A_q}(M_q) \leq \ell_{A_q}(N_q)$ for all $q \in \Spec A$ with $\dim A/q = d$ then

$$e(M/xM, A/Ax) \leq e(N/xN, A/Ax).$$

If $\ell_{A_q}(M_q) = \ell_{A_q}(N_q)$ for all $q \in \Spec A$ with $\dim A/q = d$ then

$$e(M/xM, A/Ax) = e(N/xN, A/Ax).$$

Proof. It follows from Proposition 2.2.13 in [Emerton and Gee 2014] that

$$e(M/xM, A/Ax) = \sum_q \ell_{A_q}(M_q)e(A/(q, x)),$$

where the sum is taken over all primes $q$ in the support of $M$ such that $\dim A/q = d$. The above formula implies both assertions. □

Lemma 3.13. $e(M_{\infty}(\sigma)/\varpi, R_{\infty}^{\text{inv}}(\sigma)/\varpi) \leq 2^t e(R_{\infty}^{\text{inv}}(\sigma)/\varpi)$. 

Proof. Let $T_{\infty}^{\text{inv}}(\sigma)$ be the image of $R_{\infty}^{\text{inv}}(\sigma)$ in $\End_{O}(M_{\infty}(\sigma))$. Then

$$e(T_{\infty}^{\text{inv}}(\sigma)/\varpi, R_{\infty}^{\text{inv}}(\sigma)/\varpi) \leq e(R_{\infty}^{\text{inv}}(\sigma)/\varpi).$$

If $q$ is a minimal prime of $R_{\infty}^{\text{inv}}(\sigma)$ in the support of $M_{\infty}(\sigma)$ then it follows from Lemma 3.11 that there are surjections $T_{\infty}^{\text{inv}}(\sigma) \oplus 2^t r \twoheadrightarrow M_{\infty}(\sigma)_q$. Thus $\ell(M_{\infty}(\sigma)_q) \leq 2^t r \ell(T_{\infty}^{\text{inv}}(\sigma)_q)$. The assertion follows from Lemma 3.12 applied with $x = \varpi$, $M = M_{\infty}(\sigma)$ and $N = T_{\infty}^{\text{inv}}(\sigma) \oplus 2^t r$. □

Lemma 3.14. If the support of $S_{\sigma, \psi}(U, O)_m$ meets every irreducible component of $R_{S}^{\psi, \square}(\sigma)$ then the following hold:

(i) $R_{S}^{\psi, \square}(\sigma)[[x_1, \ldots, x_g]] \twoheadrightarrow R'_{\infty}(\sigma)$ is an isomorphism.

(ii) $R_{\infty}^{\text{inv}}(\sigma)$ is reduced, equidimensional and $O$-flat.

(iii) $R_{\infty}(\sigma)$ is reduced, equidimensional and $O$-flat.

(iv) The support of $M_{\infty}(\sigma)$ meets every irreducible component of $R_{\infty}(\sigma)$.

(v) $2^t e(R_{S}^{\psi, \square}(\sigma)/\varpi, R_{\infty}^{\text{inv}}(\sigma)/\varpi) = e(M_{\infty}(\sigma)/\varpi, R_{\infty}^{\text{inv}}(\sigma)/\varpi)$.

Proof. Since $R_{S}^{\psi, \square}(\sigma)[[x_1, \ldots, x_g]]$ is reduced and equidimensional and has the same dimension as $R'_{\infty}(\sigma)$, to prove (i) it is enough to show that $R'_{\infty}(\sigma)_q \neq 0$ for every irreducible component $V(q)$ of $\Spec R_{S}^{\psi, \square}(\sigma)[[x_1, \ldots, x_g]]$. Since the diagram

$$R_{S}^{\psi, \square}(\sigma)[[x_1, \ldots, x_g]] \twoheadrightarrow R_{\infty}(\sigma)$$

$$R_{S}^{\psi, \square}(\sigma) \twoheadrightarrow R_{F,S}^{\psi}(\sigma)$$


commutes and the support of $S_{\sigma, \psi}(U, O)_m$ meets every irreducible component of $\text{Spec } S^\psi_\square$, $V(q)$ will contain a maximal ideal $n_{S}$ of $R^\psi_\square(\sigma)[[x_1, \ldots, x_g]][1/2]$, which lies in the support of $S_{\sigma, \psi}(U, O)_m$. It follows from the proof of Lemma 3.8 that (36) is an isomorphism in this case. Thus $R^\psi_\square(\sigma)[1/2] \neq 0$.

From part (i) we deduce that $R^\prime_\square(\sigma)$ is reduced, equidimensional and $O$-flat. It follows from (33) that the same holds for $R^{\text{inv}}_\square(\sigma)$. Since $R^\square_\square(\sigma)$ is a free $R^{\text{inv}}_\square(\sigma)$-module by Lemma 3.3, it is $O$-flat. Hence, it is enough to show that $R^\square_\square(\sigma)[1/2]$ is reduced and equidimensional. It follows from Lemma 3.3 that $R^\square_\square(\sigma)[1/2]$ is étale over $R^{\text{inv}}_\square(\sigma)[1/2]$, which implies the assertion. We also note that it follows from (i) that the inequality in Lemma 3.6 is an equality, and (33) implies that

$$e(R^{\text{inv}}_\square(\sigma)/\sigma) = e(R^\psi_\square/\sigma).$$

It follows from our assumption that the support of $M^\square(\sigma)$ meets every irreducible component of $R^\square_\square(\sigma)[[x_1, \ldots, x_g]]$. Part (i) and (33) imply that the support of $M^\square(\sigma)$ meets every irreducible component of $R^{\text{inv}}_\square(\sigma)$. It follows from Lemma 3.4 that the group $(\mathbb{G}_m[2])^I(\mathcal{O})$ acts transitively on the set of irreducible components of $R^\square_\square(\sigma)$ lying above a given irreducible component of $R^{\text{inv}}_\square(\sigma)$. Thus for part (iii) it is enough to show that the support of $M^\square(\sigma)$ in $\text{Spec } R^\square_\square(\sigma)$ is stable under the action of $(\mathbb{G}_m[2])^I(\mathcal{O})$. This is given by (P5) and can be proved in the same way as [Khare and Wintenberger 2009b, Lemma 9.6].

Let $V(q)$ be an irreducible component of $\text{Spec } R^\square_\square(\sigma)$. It follows from (iii) that the localization $R^\square_\square(\sigma)_q$ is a reduced artinian ring, and hence is equal to the quotient field $k(q)$. Thus $M^\square(\sigma)_q \cong M^\square(\sigma) \otimes_{R^\square_\square(\sigma)} k(q)$. It follows from Lemma 3.10 that $M^\square(\sigma)_q$ has length $r$ as an $R^\square_\square(\sigma)_q$-module. By part (iv) $M^\square(\sigma)$ is supported on every irreducible component of $R^\square_\square(\sigma)$, and thus the cycle of $M^\square(\sigma)$ is equal to $r$ times the cycle of $R^\square_\square(\sigma)$. Since both are $O$-torsion-free, we deduce that the cycle of $M^\square(\sigma)/\sigma$ is equal to $r$ times the cycle of $R^\square_\square(\sigma)/\sigma$, which implies that

$$e(M^\square(\sigma)/\sigma, R^{\text{inv}}_\square(\sigma)/\sigma) = re(R^\square_\square(\sigma)/\sigma, R^{\text{inv}}_\square(\sigma)/\sigma) = 2^r e(R^{\text{inv}}_\square(\sigma)/\sigma).$$

Part (v) follows from (39) and (40).

Proposition 3.15. For some $s \geq 0$ there is an isomorphism of $R^\psi_\square$-algebras

$$R^\psi_\square \cong R^\psi_\square[[x_1, \ldots, x_{s+|S|-1}]]/(f_1, \ldots, f_s).$$

Proof. The assertion follows from the proof of [Khare and Wintenberger 2009b, Proposition 4.5], where $s = \dim_k H^1_{L_+} (S, (\text{Ad}^0)^\psi(1))$ in the notation of that paper; see their Lemma 4.6 and the displayed equation above it.

Corollary 3.16. For some $s \geq 0$ there is an isomorphism of $R^\psi_\square(\sigma)$-algebras

$$R^\psi_\square(\sigma) \cong R^\psi_\square(\sigma)[[x_1, \ldots, x_{s+|S|-1}]]/(f_1, \ldots, f_s).$$
In particular, \( \dim R_{F,S}^{\psi,\square}(\sigma) \geq 4|S| \) and \( \dim R_{F,S}^{\psi}(\sigma) \geq 1. \)

**Proof.** Since
\[
R_{F,S}^{\psi,\square}(\sigma) \cong R_{F,S}^{\psi,\square} \otimes_{R_S^{\psi,\square}} R_S^{\psi,\square}(\sigma)
\]
the assertion follows from Proposition 3.15. Since \( \dim R_{S}^{\psi,\square}(\sigma) = 3|S| + 1 \) by (32), the isomorphism implies that
\[
\dim R_{F,S}^{\psi,\square}(\sigma) \geq 3|S| + 1 + s + |S| - 1 - s = 4|S|.
\]
Since \( R_{F,S}^{\psi,\square}(\sigma) \) is formally smooth over \( R_{F,S}^{\psi}(\sigma) \) of relative dimension \( 4|S| - 1 \), we conclude that \( \dim R_{F,S}^{\psi}(\sigma) \geq 1. \)

**Proposition 3.17.** If \( S_{\sigma,\psi}(U, \mathcal{O})_m \neq 0 \) then the following are equivalent:

1. \( 2^l r e(\mathcal{R}_{S}^{\psi,\square}(\sigma)/\mathcal{O}) = e(M_{\infty}(\sigma)/\mathcal{O}, R_{\mathcal{O}}^{\text{inv}}(\sigma)/\mathcal{O}). \)
2. \( 2^l r e(\mathcal{R}_{S}^{\psi,\square}(\sigma)/\mathcal{O}) \leq e(M_{\infty}(\sigma)/\mathcal{O}, R_{\mathcal{O}}^{\text{inv}}(\sigma)/\mathcal{O}). \)
3. the support of \( M_{\infty}(\sigma) \) meets every irreducible component of \( R_{\infty}(\sigma). \)
4. \( R_{F,S}^{\psi}(\sigma) \) is a finitely generated \( \mathcal{O} \)-module of rank at least 1 and \( S_{\sigma,\psi}(U, \mathcal{O})_n \neq 0 \) for all \( n \in m \)-Spec \( R_{F,S}^{\psi}(\sigma)[\frac{1}{2}]. \)

In this case any representation \( \rho : G_{F,S} \rightarrow \text{GL}_2(\mathcal{O}) \) corresponding to a maximal ideal of \( R_{F,S}^{\psi}(\sigma)[\frac{1}{2}] \) is modular.

**Proof.** Lemmas 3.6 and 3.13 and (33) imply that
\[
eq (M_{\infty}(\sigma)/\mathcal{O}, R_{\mathcal{O}}^{\text{inv}}(\sigma)/\mathcal{O}). \]
Thus (a) is equivalent to (b). Moreover, if (a) holds then the inequalities in the lemmas cited above have to be equalities. Since \( R_{S}^{\psi,\square}(\sigma) \) is reduced and \( \mathcal{O} \)-torsion-free, we deduce that \( R_{\mathcal{O}}'(\sigma) \cong R_{S}^{\psi,\square}(\sigma)[[x_1, \ldots, x_g]]. \) Hence, \( R_{\mathcal{O}}'(\sigma) \) is reduced, equidimensional and \( \mathcal{O} \)-torsion-free. The isomorphism (33) implies that the same holds for \( R_{\mathcal{O}}^{\text{inv}}(\sigma) \), which implies that \( R_{\mathcal{O}}(\sigma) \) is reduced, equidimensional, and \( \mathcal{O} \)-torsion-free; see the proof of Lemma 3.14. Since we have assumed (a), we have
\[
2^l r e(\mathcal{R}_{\mathcal{O}}^{\text{inv}}(\sigma)/\mathcal{O}) = e(M_{\infty}(\sigma)/\mathcal{O}, R_{\mathcal{O}}^{\text{inv}}(\sigma)/\mathcal{O}). \]
Let \( V(q_1), \ldots, V(q_m) \) be the irreducible components of the support of \( M_{\infty}(\sigma) \) in \( \text{Spec } R_{\mathcal{O}}(\sigma) \). Since \( R_{\mathcal{O}}(\sigma) \) is reduced, if \( V(q) \) is an irreducible component of \( \text{Spec } R_{\mathcal{O}}(\sigma) \) then \( \ell(R_{\mathcal{O}}(\sigma)_{q}) = 1. \) It follows from Lemma 3.10 that if \( V(q) \) is an irreducible component of \( \text{Spec } R_{\mathcal{O}}(\sigma) \) in the support of \( M_{\infty}(\sigma) \) then \( \ell(M_{\infty}(\sigma)_{q}) = r. \)
It follows from (38) that
\[ e(M_{\infty}(\sigma)/\varpi, R_{\infty}^{\text{inv}}(\sigma)/\varpi) = r \sum_{i=1}^{m} e(R_{\infty}(\sigma)/(\varpi, q_i), R_{\infty}^{\text{inv}}(\sigma)/\varpi), \]  
\[ e(R_{\infty}(\sigma)/\varpi, R_{\infty}^{\text{inv}}(\sigma)/\varpi) = \sum_{q} e(R_{\infty}(\sigma)/(\varpi, q), R_{\infty}^{\text{inv}}(\sigma)/\varpi), \]  
where the last sum is taken over all the irreducible components \( V(q) \). Since \( e(R_{\infty}(\sigma)/(\varpi, q), R_{\infty}^{\text{inv}}(\sigma)/\varpi) \neq 0 \) we deduce from (42)–(44) that (b) implies (c). We have
\[ R_{\infty}(\sigma)/(y_1, \ldots, y_{h+j}) \cong R_{F,S}^\psi(\sigma), \]
\[ M_{\infty}(\sigma)/(y_1, \ldots, y_{h+j})M_{\infty}(\sigma) \cong S_{\sigma,\psi}(U, \mathcal{O})_m. \]
Thus, if \( M_{\infty}(\sigma) \) is supported on the whole of \( \text{Spec} R_{\infty}(\sigma) \) then \( S_{\sigma,\psi}(U, \mathcal{O})_m \) is supported on the whole of \( \text{Spec} R_{F,S}^\psi(\sigma) \). Since \( S_{\sigma,\psi}(U, \mathcal{O})_m \) is a free \( \mathcal{O} \)-module of finite rank, we deduce that (c) implies (d).

If (d) holds then it follows from Corollary 3.16 that \( f_1, \ldots, f_s, \varpi \) is a part of a system of parameters of \( R_{S,\square}(\sigma)[x_1, \ldots, x_{s+|S|-1}] \), and Lemma 3.9 implies that every irreducible component of that ring contains a closed point of \( R_{F,S}^\psi(\sigma)[\frac{1}{2}] \). Since every such component is of the form \( q[x_1, \ldots, x_{s+|S|-1}] \), we deduce that every irreducible component of \( R_{S,\square}(\sigma) \) contains a closed point of \( R_{F,S}^\psi(\sigma)[\frac{1}{2}] \). It follows from the second part of (d) that the support of \( S_{\sigma,\psi}(U, \mathcal{O})_m \) meets every irreducible component of \( R_{S,\square}(\sigma) \). It follows from Lemma 3.14 that (d) implies (a). Since \( S_{\sigma,\psi}(U, \mathcal{O})[\frac{1}{2}] \) is a finite-dimensional \( L \)-vector space, the last assertion is a direct consequence of (d).

3D. Small weights. Let \( \tilde{1} \) be the trivial representation of \( \text{GL}_2(\mathbb{Z}_2) \) on a free \( \mathcal{O} \)-module of rank 1. We let \( \tilde{\sigma} \) be the space of functions \( f : \mathbb{P}^1(\mathbb{F}_2) \rightarrow \mathcal{O} \) such that \( \sum_{x \in \mathbb{P}^1(\mathbb{F}_2)} f(x) = 0 \) equipped with the natural action of \( \text{GL}_2(\mathbb{Z}_2) \). The reduction of \( \tilde{1} \) modulo \( \tilde{\sigma} \) is the trivial representation, the reduction of \( \tilde{\sigma} \) is isomorphic to \( k^2 \), which we will also denote by \( st \). These are the only smooth irreducible \( k \)-representations of \( \text{GL}_2(\mathbb{Z}_2) \).

The purpose of this subsection is to verify that the equivalent conditions of Proposition 3.17 hold when, for all \( v \mid 2 \), \( \sigma_v \) is either \( \tilde{1} \) or \( \tilde{\sigma} \), under the assumption that \( \tilde{\rho}|_{G_{F_v}} \) does not have scalar semisimplification at any place \( v \mid 2 \). If \( \sigma \) is the trivial representation then the result will follow from the modularity lifting theorem of [Khare and Wintenberger 2009b; Kisin 2009b]. In the general case, our assumption implies that any semistable lift of \( \tilde{\rho}|_{G_{F_v}} \) with Hodge–Tate weights \((0, 1)\) is crystalline (see Corollary 2.38). This implies that \( S_{1,\psi}(U, \mathcal{O})_m \) and \( S_{\sigma,\psi}(U, \mathcal{O})_m \) and \( R_{F,S}^\psi(\tilde{1}) \) and \( R_{F,S}^\psi(\sigma) \) coincide.

If \( p > 2 \), the results of this section are proved in [Gee 2011] by a characteristic-\( p \) argument.
Proposition 3.18. Assume that $\psi$ is trivial on $U \cap (\mathbb{A}_F^f)^{\times}$, $\sigma_v = \tilde{1}$ for all $v \mid 2$ and $\tilde{\rho}|_{G_v}$ does not have scalar semisimplification for any $v \mid 2$. Then $R_{F,S}^\psi(\sigma)$ is a finite $\mathcal{O}$-module of rank at least 1.

Proof. It follows from Lemma 2.2 in [Taylor 2003] that there is a finite solvable, totally real extension $F'$ of $F$ such that, for all places $w$ of $F'$ above a place $v \in S$, we have $F'_w = F_v$, except if $v \mid 2$ and $\tilde{\rho}|_{G_v}$ is unramified, in which case $F'_w$ is an unramified extension of $\mathbb{Q}_2$ and $\tilde{\rho}|_{G_{F'_w}}$ is trivial. Let $S'$ be the places of $F'$ above the places $S$ of $F$. By changing $F$ by $F'$ we are in position to apply Proposition 9.3 of [Khare and Wintenberger 2009b], part (II) of which says that the ring $R_{F',S'}^\psi(\sigma)$ is a finite $\mathcal{O}$-module. We now argue as in the last paragraph of the proof of Theorem 10.1 of [Khare and Wintenberger 2009b]. The restriction to $G_{F,U}$ induces a map between the deformation functors and hence a homomorphism $R_{F,S}^\psi(\sigma) \to R_{F,S}^{\psi'}(\sigma)$. Let $\rho_{F,S}^\psi: G_{F,S} \to \text{GL}_2(R_{F,S}^\psi(\sigma))$ be the universal deformation. Since $R_{F,S}^\psi(\sigma)/\sigma$ is finite, the image of $G_{F,S}$ in $\text{GL}_2(R_{F,S}^\psi(\sigma)/\sigma)$ under $\rho_{F,S}^\psi$ is a finite group. Since $F'/F$ is finite the image of $G_{F,S}$ in $\text{GL}_2(R_{F,S}^\psi(\sigma)/\sigma)$ is a finite group. Lemma 3.6 in [Khare and Wintenberger 2009a] implies that $R_{F,S}^\psi(\sigma)/\sigma$ is finite. Since dim $R_{F,S}^\psi(\sigma) \geq 1$ by Corollary 3.16, we conclude that dim $R_{F,S}^\psi(\sigma) = 1$ and $\sigma$ is a system of parameters for $R_{F,S}^\psi(\sigma)$, which implies that $R_{F,S}^\psi(\sigma)$ is a finite $\mathcal{O}$-module of rank at least 1.

□

Corollary 3.19. Assume that $\psi$ is trivial on $U \cap (\mathbb{A}_F^f)^{\times}$, $\sigma_v = \tilde{1}$ for all $v \mid 2$ and $\tilde{\rho}|_{G_v}$ does not have scalar semisimplification for any $v \mid 2$. If $S_{\sigma,\psi}(U, \mathcal{O})_m \neq 0$ then the equivalent conditions of Proposition 3.17 hold.

Proof. Since $S_{\sigma,\psi}(U, \mathcal{O})_m$ is nonzero and $\mathcal{O}$-torsion-free, there is a maximal ideal $\frak{n}$ of $R_{F,S}^\psi[\frac{1}{2}]$ such that $S_{\sigma,\psi}(U, \mathcal{O})_n \neq 0$. This implies that $\tilde{\rho}$ satisfies hypotheses (α) and (β) made in Section 8.2 of [Khare and Wintenberger 2009b].

Let $\frak{n}$ be any maximal ideal of $R_{F,S}^\psi[\frac{1}{2}]$, and let $\rho_{\frak{n}}$ be the corresponding representation of $G_{F,S}$. It follows from Theorem 9.7 in [Khare and Wintenberger 2009b] or Theorem 3.3.5 of [Kisin 2009b] that there is a Hilbert eigenform $f$ over $F$ such that $\rho_{\frak{n}} \cong \rho_f$. Let $\pi = \bigotimes'_v \pi_v$ be the corresponding automorphic representation of $\text{GL}_2(\mathbb{A}_F^f)$. If $v$ is a finite place, where $D$ ramifies, then, because of the way we have set up our deformation problem, $\rho_{\frak{n}}|_{G_{F_v}}$ is isomorphic to $(\gamma_v \otimes_{\psi, \text{cyc}} \ast, 0)$, where $\gamma_v$ is an unramified character. The restriction of the 2-adic cyclotomic character to $G_{F_v}$ is an unramified character which sends the arithmetic Frobenius to $q_v \in \mathbb{Z}_2^\times$. Since $\rho_{\frak{n}}$ arises from a Hilbert modular form, the representation $\rho_{\frak{n}}|_{G_{F_v}}$ is nonsplit, and this implies that $\pi_v$ is a twist of the Steinberg representation by an unramified character, at all $v$, where $D$ is ramified. By Jacquet–Langlands correspondence there is an eigenform $f^D \in S_{\sigma,\psi}(U, \mathcal{O})_m$
with the same Hecke eigenvalues as \( f \). This implies that \( S_{\sigma, \psi}(U, \mathcal{O})_m \) is supported on \( n \). Proposition 3.18 implies that part (d) of Proposition 3.17 holds.

**Lemma 3.20.** Fix a place \( w \) of \( F \) above 2. Let \( \sigma \) and \( \sigma' \) be such that for all \( v \mid 2, \ v \neq w \), we have \( \sigma_v = \sigma'_v \), which is equal to either \( \overline{1} \) or \( \overline{\text{s}t} \), and \( \sigma_w = \overline{1} \) and \( \sigma'_w = \overline{\text{s}t} \). Assume that \( \psi \) is trivial on \( U \cap (\mathbb{A}_F^F)^{\times} \), and \( \tilde{\rho}|_{G_{F_w}} \) does not have scalar semisimplification. Then the rings \( R_{F,S}^\psi(\sigma) \) and \( R_{F,S}^\psi(\sigma') \) are equal. Moreover, if \( \mathfrak{n} \) is a maximal ideal of \( R_{F,S}^\psi(\sigma)[\frac{1}{2}] \) then \( S_{\sigma, \psi}(U, \mathcal{O})_m \) is supported on \( n \) if and only if \( S_{\sigma', \psi}(U, \mathcal{O})_m \) is supported on \( n \).

**Proof.** The ring \( R_{F,\square}^\psi(\overline{1}) \) parametrizes crystalline lifts of \( \tilde{\rho}|_{G_{F_w}} \) with Hodge–Tate weights \((0, 1)\). The ring \( R_{F,\square}^\psi(\overline{\text{s}t}) \) parametrizes semistable lifts of \( \tilde{\rho}|_{G_{F_w}} \) with Hodge–Tate weights \((0, 1)\). Since both rings are reduced and \( \mathcal{O}\)-torsion-free, we have a surjection \( R_{F,\square}^\psi(\overline{\text{s}t}) \twoheadrightarrow R_{F,\square}^\psi(\overline{1}) \). The assumption that \( \tilde{\rho}|_{G_{F_w}} \) does not have scalar semisimplification implies that every such semistable lift is automatically crystalline, hence the map is an isomorphism. This implies that the global deformation rings are equal; see Corollary 2.38.

We will deduce the second assertion from the Jacquet–Langlands correspondence and the compatibility of local and global Langlands correspondence. Let \( \tau \) be either \( \sigma \) or \( \sigma' \). We fix an isomorphism \( i : \overline{\mathbb{Q}}_p \cong \mathbb{C} \), let \( \tau_\mathbb{C} = \tau \otimes_{\mathcal{O}} \mathbb{C} \) and let \( \tau_{\mathbb{C}}^\ast \) be the \( \mathbb{C} \)-linear dual of \( \tau \). Since \( U \cap (\mathbb{A}_F^F)^{\times} \) acts trivially on \( \tau \) by assumption, we may consider \( \tau_{\mathbb{C}}^\ast \) as a representation of \( U(\mathbb{A}_F^F)^{\times} \), on which \( (\mathbb{A}_F^F)^{\times} \) acts by \( \psi \). Let \( U' = \prod_v U'_v \) be an open subgroup of \( U \) such that \( U'_v = U_v \) if \( v \mid 2 \) and \( U'_v = \{ g \in U_v : g \equiv 1 \pmod{2} \} \) for all \( v \mid 2 \). Then \( U' \) acts trivially on \( \tau \). Let \( C^\infty(D^\times \setminus (D \otimes_F \mathbb{A}_F)^{\times}/U') \) be the space of smooth \( \mathbb{C} \)-valued functions on \( D^\times \setminus (D \otimes_F \mathbb{A}_F)^{\times} \) which are invariant under \( U' \). Since \( U' \) is a normal subgroup of \( U \), \( U \) acts on this space by right translations. It follows from [Kisin 2009c, §3.1.14; Taylor 2006, Lemma 1.3] that we have an isomorphism

\[
S_{\tau, \psi}(U, \mathcal{O}) \otimes_{\mathcal{O}} \mathbb{C} \cong \text{Hom}_{U(\mathbb{A}_F^F)^{\times}}(\tau, C^\infty(D^\times \setminus (D \otimes_F \mathbb{A}_F)^{\times}/U'D^\times)).
\]

This isomorphism is equivariant for the Hecke operators at \( v \not\in S \). The action of \( R_{F,S}^\psi(\tau) \) on \( S_{\tau, \psi}(U, \mathcal{O})_m \) factors through the action of the Hecke algebra \( \mathbb{T}_{\tau, \psi}(U) \). Let \( \mathfrak{n} \) be a maximal ideal of \( \mathbb{T}(U)_{\tau, \psi}[\frac{1}{2}] \). The isomorphism above implies that \( S_{\tau, \psi}(U, \mathcal{O})_m \) is nonzero if and only if there is an automorphic form

\[
f^D \in C^\infty(D^\times \setminus (D \otimes_F \mathbb{A}_F)^{\times}/U'D^\times),
\]

on which the Hecke operators for \( v \not\in S \) act by the eigenvalues given by the map \( \mathbb{T}_{\tau, \psi}(U) \to \kappa(n) \rightarrow \mathbb{C} \). Additionally, \( \text{Hom}_{U(\mathbb{A}_F^F)^{\times}}(\tau_{\mathbb{C}}^\ast, \pi) \neq 0 \), where \( \pi = \otimes_v \pi_v \) is the automorphic representation corresponding to \( f^D \).

If \( S_{\sigma, \psi}(U, \mathcal{O})_n \) is nonzero then the above implies that \( \text{Hom}_{U_w}(\mathbb{1}, \pi_w) \neq 0 \), which implies that \( \pi_w \) is an unramified principal series representation, which implies that
Hom\(_{U_w}(\tilde{\text{st}}, \pi_w) \neq 0\). Since \(\sigma_v = \sigma'_v\) for all \(v \neq w\), we conclude that \(S_{\sigma'_v, \psi}(U, \mathcal{O})_n\) is nonzero.

If \(S_{\sigma'_v, \psi}(U, \mathcal{O})_n\) is nonzero then the same argument shows that \(\text{Hom}_{U_w}(\tilde{\text{st}}, \pi_w) \neq 0\), which implies that \(\pi_w\) is either an unramified principal series representation, in which case \(\text{Hom}_{U_w}(1, \pi_w) \neq 0\) and thus \(S_{\sigma, \psi}(U, \mathcal{O})_n \neq 0\), or \(\pi_w\) is a special series. We would like to rule the last case out. By Jacquet–Langlands correspondence the functor in (45) is exact, this induces a filtration on \(M\) such that, for \(k \in \mathbb{Z}\), following [Kisin 2009a, §2.2.5] we fix a \(\mathcal{O}_n\)-module; therefore, the functor \(\otimes\) is exact. We note that since \(R\) is semistable noncrystalline. However, this cannot happen, as explained by reducing \(\sigma, \psi\) from (30) or (31). Let \(\tilde{\rho}|_{G_w}\) be the representation of \(G_{F,S}\) corresponding to the maximal ideal \(\mathfrak{n}\) of \(R_{F,S}^\psi[\frac{1}{2}]\). By the compatibility of local and global Langlands correspondence, if \(\pi'_w\) is special then \(\rho|_{G_w}\) is semistable noncrystalline. However, this cannot happen, as explained above.

**Corollary 3.21.** Assume that \(\psi\) is trivial on \(U \cap (\mathbb{A}_F^n)^\times\), \(\sigma_v\) is either \(\tilde{\mathbf{1}}\) or \(\tilde{\text{st}}\) for all \(v | 2\), and \(\tilde{\rho}|_{G_v}\) does not have scalar semisimplification for any \(v | 2\). If \(S_{\sigma, \psi}(U, \mathcal{O})_m \neq 0\) then the equivalent conditions of Proposition 3.17 hold.

**Proof.** If \(\sigma_v = \tilde{\mathbf{1}}\) for all \(v | 2\) then the assertion is proved in Corollary 3.19. Using this case and Lemma 3.20 we may show that part (d) of Proposition 3.17 is verified for all \(\sigma\) as above. □

### 3E. Computing Hilbert–Samuel multiplicity

Let \(\sigma = \bigotimes_{v | 2} \sigma_v\) be a continuous representation of \(U\) on a finitely generated \(\mathcal{O}\)-module \(W_{\sigma}\), where the \(\sigma_v\) are of the form (30) or (31). Let \(\psi : (\mathbb{A}_F^n)^\times / F^\times \rightarrow \mathcal{O}^\times\) be a continuous character such that \(U \cap (\mathbb{A}_F^n)^\times\) acts on \(W_{\sigma}\) by the character \(\psi\). Let \(\sigma\) be a representation of \(U\). Assume that \(U\) satisfies (29), which implies that the subgroups \(U_{Q_n}\) also satisfy (29). Hence, the functor \(\sigma \mapsto S_{\sigma, \psi}(U_{Q_n}, \mathcal{O})\) is exact. We note that since \(R_{F,S}^\psi, \square\) is formally smooth over \(R_{F,S}^\psi\), it is a flat \(R_{F,S}^\psi\)-module; therefore, the functor \(\bigotimes_{\mathcal{O}} R_{F,S}^\psi\) is exact, and so is the localization at \(m_{Q_n}\). Hence the functor

\[
\sigma \mapsto M_n(\sigma) = R_{F,S_{Q_n}}^\psi, \square \otimes_{R_{F,S_{Q_n}}^\psi} S_{\sigma, \psi}(U_{Q_n}, \mathcal{O})_{m_{Q_n}} \quad (45)
\]

is exact. Following [Kisin 2009a, §2.2.5] we fix a \(U\)-invariant filtration on \(\tilde{\sigma}\) by \(k\)-subspaces

\[0 = L_0 \subset L_1 \subset \cdots \subset L_s = W_{\sigma} \otimes_{\mathcal{O}} k\]

such that, for \(i = 0, 1, \ldots, s - 1\), \(\sigma_i := L_{i+1}/L_i\) is absolutely irreducible. Since the functor in (45) is exact, this induces a filtration on \(M_n(\sigma) \otimes_{\mathcal{O}} k\), which we denote by

\[
0 = M_{n}^0(\sigma) \subset M_{n}^1(\sigma) \subset \cdots \subset M_{n}^s(\sigma) = M_{n}(\sigma) \otimes_{\mathcal{O}} k, \quad (46)
\]
such that, for \( i = 0, 1, \ldots, s - 1 \), we have

\[
M_n^{i+1}(\sigma)/M_n^i(\sigma) \cong M_n(\sigma_i). \tag{47}
\]

Each representation \( \sigma_i \) is of the form \( \bigotimes_{v \mid 2} \sigma_{i,v} \), where \( \sigma_{i,v} \) is either the trivial representation, in which case we let \( \tilde{\sigma}_{i,v} = 1 \), or st, in which case we let \( \tilde{\sigma}_{i,v} := \tilde{st} \). We let \( \tilde{\sigma}_i := \bigotimes_{v \mid 2} \tilde{\sigma}_{i,v} \) and consider it as a representation of \( U \) by letting \( U_v \) for \( v \) not above 2 act trivially. We note that, since both \( \tilde{1} \) and \( \tilde{st} \) have trivial central character, \( U \cap (\mathbb{A}_F^f)^{\times} \) acts trivially on \( \tilde{\sigma}_i \). We choose a continuous character \( \xi : F^\times \backslash (\mathbb{A}_F^f)^{\times} \to \mathcal{O}^\times \) such that \( \psi \equiv \xi \pmod{\mathfrak{o}} \) and the restriction of \( \xi \) to \( U \cap (\mathbb{A}_F^f)^{\times} \) is trivial. For example, we could choose \( \xi \) to be a Teichmüller lift of \( \bar{\psi} \). Let

\[
M_n(\tilde{\sigma}_i) = R_{F,S_{Q_n}}^{\xi,\square} \otimes R_{F,S_{Q_n}}^{\xi,\square} S_{\tilde{\sigma}_i,\xi}(U_{Q_n}, \mathcal{O}) m_{Q_n}.
\]

The exactness of the functor in (45), used with \( \tilde{\sigma}_i \) and \( \xi \) instead of \( \sigma \) and \( \psi \), and (47) give us an isomorphism

\[
\alpha_{i,n} : M_n^{i+1}(\sigma)/M_n^i(\sigma) \cong M_n(\sigma_i) \cong M_n(\tilde{\sigma}_i) \otimes_{\mathcal{O}} k. \tag{48}
\]

The isomorphism \( \alpha_{i,n} \) is equivariant for the action of the Hecke operators outside \( S_{Q_n} \), since they act by the same formulas on all the modules. Hence (48) is an isomorphism of \( R_S^{\square}[x_1, \ldots, x_g] \)-modules. We let \( \alpha_{i,n} \) be the \( R_{F,S_{Q_n}}^{\xi,\square}(\tilde{\sigma}_i) \)-annihilator of \( M_n(\tilde{\sigma}_i) \otimes_{\mathcal{O}} k \). Since the action of \( R_S^{\square}[x_1, \ldots, x_g] \) on \( M_n(\sigma_1) \) and \( M_n(\tilde{\sigma}_i) \) factors through \( R_{F,S_{Q_n}}^{\psi,\square}(\sigma) \) and \( R_{F,S_{Q_n}}^{\xi,\square}(\tilde{\sigma}_i) \), respectively, we obtain a surjection

\[
\varphi_{i,n} : R_{F,S_{Q_n}}^{\psi,\square}(\sigma) \twoheadrightarrow R_{F,S_{Q_n}}^{\xi,\square}(\tilde{\sigma}_i)/\alpha_{i,n}. \tag{49}
\]

**Proposition 3.22.** We may patch in such a way that:

- There is an \( R_\infty(\sigma) \)-module \( M_\infty(\sigma) \) as in Section 3C.
- There is a filtration

\[
0 = M^0_\infty(\sigma) \subset M^1_\infty(\sigma) \subset \cdots \subset M^s_\infty(\sigma) = M_\infty(\sigma) \otimes_{\mathcal{O}} k
\]

by \( R_\infty(\sigma) \)-submodules.
- For each \( 1 \leq i \leq s \) there is an \( R_\infty(\tilde{\sigma}_i) \)-module \( M_\infty(\tilde{\sigma}_i) \) as in Section 3C and a surjection \( \varphi_i : R_\infty(\sigma) \twoheadrightarrow R_\infty(\tilde{\sigma}_i)/\alpha_i \), where \( \alpha_i \) is the \( R_\infty(\tilde{\sigma}_i) \)-annihilator of \( M_\infty(\tilde{\sigma}_i) \otimes_{\mathcal{O}} k \), which allows us to consider \( M_\infty(\tilde{\sigma}_i) \otimes_{\mathcal{O}} k \) as an \( R_\infty(\sigma) \)-module.
- For each \( 1 \leq i \leq s \) there is an isomorphism of \( R_\infty(\sigma) \)-modules

\[
\alpha_i : M^i_\infty(\sigma)/M^{i-1}_\infty(\sigma) \cong M_\infty(\tilde{\sigma}_i) \otimes_{\mathcal{O}} k.
\]

**Proof.** We modify the proof of [Khare and Wintenberger 2009b, Proposition 9.3], which in turn is a modification of the proof of [Kisin 2009c, Proposition 3.3.1]. Let \( \Delta(\sigma)_m := (D(\sigma)_m, L(\sigma)_m, D'(\sigma)_m) \) be the patching data of level \( m \) as in the proof
of [Khare and Wintenberger 2009b, Proposition 9.3], where $\sigma$ indicates the fixed weight and inertial type we are working with. In particular, $D(\sigma)_m$ and $D'(\sigma)_m$ are finite $R^\psi_S \boxplus (\sigma) \llbracket x_1, \ldots, x_g \rrbracket$-algebras, where $g = h + j + t - d$, and $L(\sigma)_m$ is a module over $D(\sigma)_m$ satisfying a number of conditions, listed in the proof of [Khare and Wintenberger 2009b, Proposition 9.3]. Our patching data of level $m$ consists of tuples

$$\Delta_m := (\{ \Delta(\sigma)_m \}, \{ L(\sigma)^i_m \}_{i=0}^s, \{ \Delta(\bar{\sigma}_i)_m \}_{i=1}^s, \{ \phi_{i,m} \}_{i=1}^s, \{ \alpha_{i,m} \}_{i=1}^s),$$

where $\{ L(\sigma)^i_m \}_{i=0}^s$ is a filtration of $L(\sigma)_m \otimes \mathcal{O} k$ by $D(\sigma)_m$-submodules, $\phi_{i,m} : D(\sigma)_m \twoheadrightarrow D(\bar{\sigma}_i)_m/\alpha_{i,m}$ is a surjection of $R_S \llbracket x_1, \ldots, x_g \rrbracket$-algebras, where $\alpha_{i,m}$ is the $D(\bar{\sigma}_i)_m$-annihilator of $L(\bar{\sigma}_i) \otimes \mathcal{O} k$, and $\alpha_{i,m}$ is an isomorphism of $D(\sigma)_m$-modules between $L(\sigma)^i_m / L(\sigma)^{i-1}_m$ and $L(\bar{\sigma}_i) \otimes \mathcal{O} k$, where the action of $D(\sigma)_m$ on this last module is given by $\phi_{i,m}$.

An isomorphism of patching data between $\Delta_m$ and $\Delta'_m$ is a tuple $(\beta, \{ \beta_i \}_{i=1}^s)$, where $\beta : \Delta_m(\sigma) \cong \Delta'_m(\sigma)$ and $\beta_i : \Delta_m(\bar{\sigma}_i) \cong \Delta_m(\bar{\sigma}_i)$ are isomorphisms of patching data, in the sense of [Khare and Wintenberger 2009b, Proposition 9.3], which respect the filtration and the maps $\{ \phi_{i,m} \}_{i=1}^s, \{ \alpha_{i,m} \}_{i=1}^s$. There are only finitely many isomorphism classes of patching data of level $m$, since there are only finitely many isomorphism classes of patching data of level $m$ in the sense of [Khare and Wintenberger 2009b, Proposition 9.3], and a finite $\mathcal{O}$-module can admit only finitely many filtrations and there are only finitely many maps between two finite modules.

We then proceed as in the proof of [Khare and Wintenberger 2009b, Proposition 9.3]. In particular, the integers $a, r_m, n_0$ and ideals $c_m$ and $b_n$ are those defined in [loc. cit.]. For an integer $n \geq n_0 + 1$ and for $m$ with $n \geq m \geq 3$, let $\Delta_{n,m}(\sigma) = (D(\sigma)_{n,m}, L(\sigma)_{n,m}, D'(\sigma)_{n,m})$ be the patching data of level $m$ as in the proof of [Khare and Wintenberger 2009b, Proposition 9.3]. Then

$$D(\sigma)_{n,m} = R_{n+a}(\sigma) / (c_m R_{n+a}(\sigma) + m_{R_{n+a}(\sigma)} (r_m),$$

$$L(\sigma)_{n,m} = M_{n+a}(\sigma) / c_m M_{n+a}(\sigma),$$

where $R_n(\sigma) := R_{F,S_n}(\sigma)$. We define $\Delta_{n,m}(\bar{\sigma}_i)$ analogously with $\bar{\sigma}_i$ instead of $\sigma$ and with $\xi$ instead of $\psi$. We let $L(\sigma)_{n,m}^i_{i=1}$ be the filtration obtained by reducing (46) modulo $c_m$. Similarly, we let $\{ \phi_{i,n,m} \}_{i=1}^s, \{ \alpha_{i,n,m} \}_{i=1}^s$ be the maps obtained by reducing (48) and (49) modulo $c_m$. Then

$$\Delta_{n,m} := (\{ \Delta(\sigma)_{n,m} \}, \{ L(\sigma)^i_{n,m} \}_{i=0}^s, \{ \Delta(\bar{\sigma}_i)_{n,m} \}_{i=1}^s, \{ \phi_{i,n,m} \}_{i=1}^s, \{ \alpha_{i,n,m} \}_{i=1}^s)$$

is a patching datum of level $m$ in our sense. Since there are only finitely many isomorphism classes of patching data of level $m$, after replacing the sequence

$$((R_{n+a}(\sigma), M_{n+a}(\sigma)), ((R_{n+a}(\bar{\sigma}_i), M_{n+a}(\bar{\sigma}_i)))_{i=1}^s)_{n \geq n_0 + 1}$$
by a subsequence, we may assume that, for each \( m \geq n_0 + 4 \) and all \( n \geq m \), we have \( \Delta_{m,n} = \Delta_{m,m} \). The patching data \( \Delta_{m,m} \) form a projective system; see [Kisin 2009c, Proposition 3.3.1]. We obtain the desired objects by passing to the limit. \( \square \)

We need to control the image of \( R_\infty^{inv}(\sigma) \) under \( \varphi_i \). Following [Khare and Wintenberger 2009b] we let \( \text{CNL}_O \) be the category of complete local noetherian \( \mathcal{O} \)-algebras with a fixed isomorphism of the residue field with \( k \), and whose maps are local \( \mathcal{O} \)-algebra homomorphisms. If \( A \in \text{CNL}_O \) then we let \( \text{Sp}_A : \text{CNL}_O \to \text{Sets} \) be the functor \( \text{Sp}_A(B) = \text{Hom}_{\text{CNL}_O}(A, B) \). Let \( G \) be a finite abelian group. We let \( G^* \) be the group scheme defined over \( \mathcal{O} \) such that, for every \( \mathcal{O} \)-algebra \( A \), \( G^*(A) = \text{Hom}_{\text{Groups}}(G, A^\times) \). Assume that we are given a free \( G^* \) action on \( \text{Sp}_A \). This means that, for all \( B \in \text{CNL}_O \), \( G^*(B) \) acts on \( \text{Sp}_A(B) \) without fixed points. By Proposition 2.6(1) in [Khare and Wintenberger 2009b] the quotient \( G^* \backslash \text{Sp}_A \) exists in \( \text{CNL}_O \) and is represented by \( (A^{inv}, m_A^{inv}) \in \text{CNL}_O \). Moreover, \( \text{Sp}_A \) is a \( G^* \)-torsor over \( \text{Sp}_{A^{inv}} \).

**Lemma 3.23.** Let \( (A, m_A) \) and \( (B, m_B) \) be in \( \text{CNL}_O \). Assume that \( G^* \) acts freely on \( \text{Sp}_A \) and \( \text{Sp}_B \) and we are given a \( G^* \)-equivariant closed immersion \( \text{Sp}_B \hookrightarrow \text{Sp}_A \). Then the map induces a closed immersion \( \text{Sp}_{B^{inv}} \hookrightarrow \text{Sp}_{A^{inv}} \).

**Proof.** Since \( G^* \) acts trivially on \( \text{Sp}_{A^{inv}} \), by the universal property of the quotient, the map \( \text{Sp}_B \to \text{Sp}_A \to \text{Sp}_{A^{inv}} \) factors through \( \text{Sp}_{B^{inv}} \to \text{Sp}_{A^{inv}} \). Hence, we obtain the following commutative diagram in \( \text{CNL}_O \):

\[
\begin{array}{ccc}
A^{inv} & \to & B^{inv} \\
\downarrow & & \downarrow \\
A & \to & B
\end{array}
\]

Since \( \text{Sp}_A \) is a \( G^* \)-torsor over \( \text{Sp}_{A^{inv}} \), it follows from [SGA 3 II 1970, Exposé VIII, Proposition 4.1] that \( A \) is a free \( A^{inv} \)-module of rank \( |G| \). Similarly, \( B \) is a free \( B^{inv} \)-module of rank \( |G| \). It follows from the commutative diagram that the surjection \( A \to B \) induces a surjection \( A/m_A^{inv} A \to B/m_B^{inv} B \). Since both \( k \)-vector spaces have dimension \( |G| \), the map is an isomorphism and this implies that the image of \( m_A^{inv} \) is equal to \( m_B^{inv} \). Hence, the top horizontal arrow in the diagram is surjective. \( \square \)

Let \( \text{CNL}_O^{[m]} \) be the full subcategory of \( \text{CNL}_O \) consisting of objects \( (A, m_A) \) such that \( m_A^m = 0 \). We have a truncation functor \( \text{CNL}_O \to \text{CNL}_O^{[m]}, A \mapsto A^{[m]} := A/m_A^m \). If \( A \) represents the functor \( X \), we denote by \( X^{[m]} \) the functor represented by \( A^{[m]} \). For group chunk actions, we refer the reader to [Khare and Wintenberger 2009b, §2.6].

**Lemma 3.24.** Let \( (A, m_A) \) and \( (B, m_B) \) be in \( \text{CNL}_O \). Assume that \( G^* \) acts freely on \( X := \text{Sp}_A \) and \( Y := \text{Sp}_B \) and we are given an isomorphism \( X^{[m]} \cong Y^{[m]} \) compatible with the group chunk \( (G^*)^{[m]} \)-action. If \( m \) is large enough then the image of \( m_A^{inv} A \) in \( A/m_A^m = B/m_B^m \) is equal to the image of \( m_B^{inv} B \).
Proof. Let $X^\text{inv}$ and $Y^\text{inv}$ denote the quotients of $X$ and $Y$ by $G^*$. Then we have isomorphisms

$$G^* \times X \cong X \times X^\text{inv}, \quad G^* \times Y \cong Y \times Y^\text{inv},$$

where the map is given by $(g, x) \mapsto (x, gx)$. We define $Z := X^{[m]} = Y^{[m]}$ and $C := A/m_A^m = B/m_B^m$. The restriction of the above isomorphism to $\text{CNL}_Z^{[m]}$ gives us isomorphisms

$$(G^* \times Z)^{[m]} \cong (Z \times X^\text{inv} Z)^{[m]}, \quad (G^* \times Z)^{[m]} \cong (Z \times Y^\text{inv} Z)^{[m]}.$$

Thus we have an isomorphism

$$(Z \times X^\text{inv} Z)^{[m]} \cong (Z \times Y^\text{inv} Z)^{[m]},$$

where the map is given by $(z_1, z_2) \mapsto (z_1, z_2)$. On rings this isomorphism reads $(C \otimes_{A^\text{inv}} C)^{[m]} \cong (C \otimes_{B^\text{inv}} C)^{[m]}$, $c_1 \otimes c_2 \mapsto c_1 \otimes c_2$.

Both $A/m_A^m A$ and $B/m_B^m B$ are $k$-vector spaces of dimension $|G|$. In particular, if $m > |G|$ then $m_A^m \subset m_A^1$ and $m_B^m \subset m_B^1$. So we obtain a map $C \rightarrow A/m_A^m A$. If $m > 2|G|$ then by base changing along this map, we obtain an isomorphism

$$A/m_A^m A \otimes_k A/m_A^m A \cong A/m_A^m A \otimes_{B^m} A/m_A^m A.$$

If the image of $B^m$ in $A/m_A^m A$ is not equal to $k$ then, for some $b \in B^m$, $1 \otimes b$ and $b \otimes 1$ will be linearly independent over $k$ in the left-hand side of the above isomorphism and linearly dependent in the right-hand side. This implies that the image of $B^m$ in $A/m_A^m A$ is equal to $k$. Thus $m_B^m C \subset m_A^m C$ and by symmetry we obtain the other inclusion. \hfill $\Box$

Let $G_n$ be the Galois group of the maximal abelian extension of $F$, of degree a power of 2, which is unramified outside $Q_n$ and split at primes in $S$. Let $G_{n,2} = G_n/2G_n$. It follows from [Khare and Wintenberger 2009b, Lemma 5.1(f)] that $G_{n,2} \cong (\mathbb{Z}/2\mathbb{Z})^I$. Let $G_{n,2}^*$ be the group scheme defined over $\mathcal{O}$ such that, for every $\mathcal{O}$-algebra $A$, $G_{n,2}^*(A) = \text{Hom}_{\text{Groups}}(G_{n,2}, A^\times)$. For a local artinian augmented $\mathcal{O}$-algebra $A$ and $\chi \in G_{n,2}^*(A)$, if $\rho_A$ is a $G_{F, S\mathcal{Q}_n}$-representation lifting $\bar{\rho}$ to $A$ then so is $\rho_A \otimes \chi$. Moreover, since $\chi^2$ is trivial, $\rho_A$ and $\rho_A \otimes \chi$ have the same determinant. This induces an action of $G_{n,2}^*$ on

$$\text{Spf} R_{F, S\mathcal{Q}_n}, \quad \text{Spf} R_{F, S\mathcal{Q}_n}^\psi, \quad \text{and} \quad \text{Spf} R_{F, S\mathcal{Q}_n}^{\hat{\psi}}.$$

It follows from [Khare and Wintenberger 2009b, Lemma 5.1] that this action is free. Proposition 2.6 of [Khare and Wintenberger 2009b] implies that the quotient by $G_{n,2}^*$ is represented by a complete local noetherian $\mathcal{O}$-algebra, which we will denote by $(R_{F, S\mathcal{Q}_n}^{\Box, m} m_{n, \sigma}), (R_{F, S\mathcal{Q}_n}^{\psi, \Box, m} m_{n, \sigma})$ and $(R_{F, S\mathcal{Q}_n}^{\hat{\psi}, \Box, m} m_{n, \sigma})$, respectively.
Lemma 3.25. The map

$$\text{Spf } R_{F,S_{Q_n}}^\xi,\Box(\tilde{s}_i)/\alpha_{i,n} \to \text{Spf } R_{F,S_{Q_n}}^\psi,\Box(\sigma)$$

induced by (49) is $G_{n,2}^*$-equivariant. Moreover,

$$\varphi_{i,n}(m_{n,\sigma}^\psi R_{F,S_{Q_n}}^\psi,\Box(\sigma)) = m_{n,\tilde{s}_i}^\psi R_{F,S_{Q_n}}^\xi,\Box(\tilde{s}_i)/\alpha_{i,n}.$$  

Proof. The first part follows from [Khare and Wintenberger 2009b, Lemma 9.1]; see the paragraph after the proof of Proposition 7.6 and the third paragraph of the proof of Lemma 9.6 of the same paper.

Let

$$q_\sigma : R_{F,S_{Q_n}}^\Box \to R_{F,S_{Q_n}}^\psi,\Box(\sigma) \quad \text{and} \quad q_{\tilde{s}_i} : R_{F,S_{Q_n}} \to R_{F,S_{Q_n}}^\xi,\Box(\tilde{s}_i)$$

denote the natural surjections. Since $\varphi_{i,n} \circ q_\sigma = q_{\tilde{s}_i} (\text{mod } \alpha_{n,i})$, it is enough to show that $q_\sigma (m_{n,\sigma}^\psi R_{F,S_{Q_n}}^\psi,\Box) = m_{n,\tilde{s}_i}^\psi R_{F,S_{Q_n}}^\xi,\Box(\tilde{s}_i)$ for all $\sigma$ and $\psi$ as above. This follows from Lemma 3.23. \qed

Let $m_{\sigma}^\psi$ and $m_{\tilde{s}_i}^\psi$ be the maximal ideals of $R_{\infty}^\psi(\sigma)$ and $R_{\infty}^\psi(\tilde{s}_i)$, respectively.

Proposition 3.26. The surjection $\varphi_i : R_{\infty}(\sigma) \to R_{\infty}(\tilde{s}_i)/\alpha_i$ maps $m_{\sigma}^\psi R_{\infty}(\sigma)$ onto the image of $m_{\tilde{s}_i}^\psi R_{\infty}(\tilde{s}_i)$. In particular,

$$e(M_{\infty}^i(\sigma)/M_{\infty}^{i-1}(\sigma), R_{\infty}^\psi(\sigma)/\alpha_i) = e(M_{\infty}(\tilde{s}_i) \otimes \mathbb{Q}_p k, R_{\infty}^\psi(\tilde{s}_i)/\alpha_i).$$  

(50)

Proof. If $(A, m)$ is a complete local noetherian algebra then by $A^{[r]}$ we denote the ring $A/m^r$. We will use the same notation as in the proof of the previous proposition. It is shown in the course of the proof of part (I) of [Khare and Wintenberger 2009b, Proposition 9.3] that

$$R_{\infty}(\sigma) \cong \lim_m D_{m,m}^\prime(\sigma),$$

where $D_{m,n}^\prime(\sigma) = R_{n+a}(\sigma)^{[m]}$. Moreover, it is shown that the map is $(\widehat{G}_{m}[2])^l$-equivariant by fixing an identification of $G_{n+a,2}$ with $(\mathbb{Z}/2\mathbb{Z})^l$.

For each fixed $r \geq 0$ we have

$$R_{\infty}(\sigma)^{[r]} \cong \lim_m D_{m,m}^\prime(\sigma)^{[r]}.$$ 

Hence, by choosing $m$ large enough we may assume that $R_{\infty}(\sigma)^{[r]} = D_{m,m}^\prime(\sigma)^{[r]}$ with $r \leq r'_m$. Since $(\widehat{G}_{m}[2])^l$-action on $S_{P_{\infty}(\sigma)}$ and on $S_{P_{R_{n+a}}(\sigma)}$ is free by [Khare and Wintenberger 2009b, Lemmas 5.1 and 9.4], we are in the situation of Lemma 3.24. Hence the image of $m_{\sigma}^\psi R_{\infty}(\sigma)$ in $D_{m,m}^\prime(\sigma)^{[r]}$ is equal to the image of $m_{m+a,a}^\psi R_{m+a}(\sigma)$. It follows from Lemma 3.25 that the composition

$$R_{\infty}(\sigma) \to R_{m+a}(\sigma)^{[r]} \overset{\varphi_{i,m}}{\to} (R_{m+a}(\tilde{s}_i)/\alpha_{i,m})^{[r]}$$

is $G_{n,2}^*$-equivariant.
maps $m^\text{inv}_\sigma R_\infty(\sigma)$ onto the image of $m^\text{inv}_{\tilde{\sigma}_i} R_\infty(\tilde{\sigma}_i)$. The action of $R_{m+a}(\tilde{\sigma}_i)$ on $L_{m,m}(\tilde{\sigma}_i)$ factors through $R_{m+a}(\tilde{\sigma}_i)[m^{-1}]$. Since by construction

$$\varphi_i = \lim_{m} \varphi_i,m, \quad R_\infty(\tilde{\sigma}_i) = \lim_{m} R_{m+a}(\tilde{\sigma}_i)[m^{-1}], \quad M_\infty(\tilde{\sigma}_i) = \lim_{m} L_{m,m}(\tilde{\sigma}_i),$$

we deduce that $\varphi_i$ maps $m^\text{inv}_\sigma R_\infty(\sigma)$ onto the image of $m^\text{inv}_{\tilde{\sigma}_i} R_\infty(\tilde{\sigma}_i).$ 

**Corollary 3.27.** Assume that $S_{\sigma,\psi}(U, \mathcal{O})_m \neq 0$ and that $\tilde{\rho}|_{G_{F_v}} \not\cong \begin{pmatrix} \chi & * \\ 0 & \chi \end{pmatrix}$ for $v \mid 2$ and any character $\chi : G_{F_v} \to k^\times$. Then the equivalent conditions of Proposition 3.17 hold, and any $\rho : G_{F,S} \to \text{GL}_2(\mathcal{O})$ corresponding to a maximal ideal of $R^\psi_{F,S}(\sigma)[\frac{1}{2}]$ is modular.

**Proof.** We will verify that part (b) of Proposition 3.17 holds. We first note that, since $S_{\sigma,\psi}(U, \mathcal{O})_m \neq 0$ and $U$ satisfies (29), there is an $i$ such that $S_{\tilde{\sigma}_i,\xi}(U, k)_m \neq 0$. This implies that $S_{\tilde{\sigma}_i,\xi}(U, \mathcal{O})_m \neq 0$, and it follows from Lemma 3.20 that $S_{\tilde{\sigma}_i,\xi}(U, \mathcal{O})_m \neq 0$ for all $1 \leq i \leq s$ and $S_{1,\xi}(U, \mathcal{O})_m \neq 0$. In particular, the rings $R^\psi_S(\tilde{\sigma}_i)$ are nonzero and equal to $R^\psi_S(\tilde{\sigma}_i)$. Corollary 3.21 implies that for all $1 \leq i \leq s$ the equality

$$2^l r e\left(R^\psi_S(\tilde{\sigma}_i)/\mathcal{O}\right) = e\left(M_\infty(\tilde{\sigma}_i)/\mathcal{O}, R^\text{inv}_\infty(\tilde{\sigma}_i)/\mathcal{O}\right)$$

holds. Since the Hilbert–Samuel multiplicity is additive in short exact sequences, we have

$$e\left(M_\infty(\sigma)/\mathcal{O}, R^\text{inv}_\infty(\sigma)/\mathcal{O}\right) = \sum_{i=1}^{s} e\left(M^i_\infty(\sigma)/M^{i-1}_\infty(\sigma), R^\text{inv}_\infty(\sigma)/\mathcal{O}\right).$$

Proposition 3.26 implies that for all $1 \leq i \leq s$ we have

$$e\left(M^i_\infty(\sigma)/M^{i-1}_\infty(\sigma), R^\text{inv}_\infty(\sigma)/\mathcal{O}\right) = e\left(M_\infty(\tilde{\sigma}_i)/\mathcal{O}, R^\text{inv}_\infty(\tilde{\sigma}_i)/\mathcal{O}\right).$$

Thus

$$e\left(M_\infty(\sigma)/\mathcal{O}, R^\text{inv}_\infty(\sigma)/\mathcal{O}\right) = 2^l r \sum_{i=1}^{s} e\left(R^\psi_S(\tilde{\sigma}_i)/\mathcal{O}\right).$$

Thus to verify part (b) of Proposition 3.17 it is enough to show that

$$e\left(R^\psi_S(\sigma)/\mathcal{O}\right) \leq \sum_{i=1}^{s} e\left(R^\psi_S(\tilde{\sigma}_i)/\mathcal{O}\right).$$

If $A$ and $B$ are complete local $\kappa$-algebras with residue field $\kappa$ then it is shown in [Kisin 2009a, Proposition 1.3.8] that $e(A \hat{\otimes}_\kappa B) = e(A)e(B)$. Since $\psi$ is congruent to $\xi$ modulo $\mathcal{O}$, inequality (55) reduces to the following inequality on Hilbert–Samuel multiplicities of potentially semistable rings at all $v \mid 2$:

$$e\left(R^\psi_v(\sigma)/\mathcal{O}\right) \leq \sum_{i=1}^{s_v} e\left(R^\psi_v(\tilde{\sigma}_{v,i})/\mathcal{O}\right).$$
Here the \( \sigma_{v,i} \) are irreducible \( k \)-representation of \( \GL_2(F_2) \) which appear as graded pieces of a \( \GL_2(\mathbb{Z}_2) \)-invariant filtration on \( \sigma_v \otimes_{\mathcal{O}} k \). Inequality (56) is proved in the local part of the paper; see Remark 2.39. \( \square \)

3F. Modularity lifting. Let \( F \) be a totally real field in which 2 splits completely.

**Definition 3.28.** An allowable base change is a totally real solvable extension \( F' \) of \( F \) such that 2 splits completely in \( F' \).

**Lemma 3.29.** Assume that \( [F : \mathbb{Q}] \) is even. Let \( \bar{\rho} : G_F \to \GL_2(k) \) be a continuous absolutely irreducible representation. If there is a Hilbert eigenform \( f \) such that \( \bar{\rho} \equiv \bar{\rho}_f \) then there is a Hilbert eigenform \( g \) of parallel weight 2 such that \( \bar{\rho} \equiv \bar{\rho}_g \) and at \( v | 2 \) the corresponding representation \( \pi_v \) of \( \GL_2(F_v) \) is either an unramified principal series or a twist of Steinberg representation by an unramified character. Moreover, if \( \bar{\rho}|_{G_{F_v}} \not\cong (\chi \, \ast \, 0) \) for all \( v | 2 \) and any character \( \chi : G_{F_v} \to k^* \) then we may assume that \( \pi_v \) is an unramified principal series representation for all \( v | 2 \).

**Proof.** Let \( D \) be the totally definite quaternion algebra with center \( F \) split at all the finite places. Let \( f^D \in S_{\tau,\psi}(U, \mathcal{O}) \) be the eigenform on \( D \) associated to \( f \) by the Jacquet–Langlands correspondence, where \( U = \prod_v U_v \) is a compact open subgroup of \( (D \otimes_F \mathbb{A}_F)^\times \) such that \( U_v = \GL_2(O_{F_v}) \) for all \( v | 2 \), and \( U \) is sufficiently small, so that (29) holds, and \( \tau = \bigotimes_v \tau_v \) is a locally algebraic representation of \( U \). Let \( m \) be the maximal ideal of the Hecke algebra \( \mathcal{T}_{\univ}^u \) corresponding to \( \bar{\rho} \). Then \( f^D \in S_{\tau,\psi}(U, \mathcal{O})_m \), and hence \( S_{\tau,\psi}(U, \mathcal{O})_m \) is nonzero.

Let \( \bar{\tau} \) denote the reduction of a \( U \)-invariant lattice in \( \tau \), and let \( \bar{\psi} \) denote \( \psi \) modulo \( \varsigma \). Since \( U \) satisfies (29) the functor \( \sigma \mapsto S_{\sigma,\psi}(U, \mathcal{O}) \) is exact. The localization functor is also exact. Hence there is an irreducible subquotient \( \sigma \) of \( \bar{\tau} \) such that \( S_{\sigma,\psi}(U, k)_m \) is nonzero. Such a \( \sigma \) is of the form \( \bigotimes_{v|2} \sigma_v \), where \( \sigma_v \) is a representation of \( \GL_2(F_2) \). Thus \( \sigma_v \) is either trivial, in which case we let \( \bar{\sigma}_v = \bar{1} \), or \( k^2 \), in which case we let \( \bar{\sigma}_v = \bar{s} \tilde{t} \). Then the reduction of \( \bar{\sigma}_v \) modulo \( \varsigma_v \) is isomorphic to \( \sigma_v \) and \( F_v^\times \cap U_v \) acts trivially on \( \bar{\sigma}_v \). Let \( \bar{\sigma} := \bigotimes_{v|2} \bar{\sigma}_v \). Choose a lift \( \xi : (\mathbb{A}_F^f)^\times / F^\times \to \mathcal{O}^\times \) of \( \bar{\psi} \), which is trivial on \( U \cap (\mathbb{A}_F^f)^\times \). The exactness of the functor \( \sigma \mapsto S_{\sigma,\xi}(U, \mathcal{O}) \) implies that \( S_{\bar{\sigma},\xi}(U, \mathcal{O})_m \) is nonzero, since its reduction modulo \( \varsigma \) is equal to \( S_{\bar{\sigma},\xi}(U, k)_m \). We may take any eigenform \( g^D \in S_{\bar{\sigma},\bar{\psi}}(U, \mathcal{O})_m \) and then using Jacquet–Langlands transfer it to a Hilbert modular form, which will have the prescribed properties. The last part follows from Lemma 3.20. \( \square \)

**Theorem 3.30.** Let \( F \) be a totally real field where 2 is totally split, and let
\[
\rho : G_{F,S} \to \GL_2(\mathcal{O})
\]
be a continuous representation. Suppose:
Then there is an admissible base change \( \gamma \) such that

\[
\operatorname{det} \rho \prod S \quad \text{deduce that}
\]

for all particular, of \( U \) correspondence. Then \( \rho \) is totally odd.

(iv) If \( v \mid 2 \) then \( \bar{\rho}|_{G_{\mathbb{F}_v}} \not\cong (\chi \circ \gamma_v)^* \), for any character \( \chi : G_{\mathbb{F}_v} \to k^\times \).

Then \( \rho \) is modular.

**Proof.** Let \( \psi = \chi_{\text{cyc}}^{-1} \operatorname{det} \rho \), where \( \chi_{\text{cyc}} \) is the 2-adic cyclotomic character. By solvable base change it is enough to prove the assertion for the restriction of \( \rho \) to \( G_{F'} \), where \( F' \) is a totally real solvable extension of \( F \). Using Lemma 2.2 of [Taylor 2003] we may find an allowable base change \( F' \) of \( F \) such that \( [F' : \mathbb{Q}] \) is even and \( \bar{\rho}|_{G_{F'}} \) is unramified outside places above 2. We may further assume that if \( \rho \) is ramified at \( v \mid 2 \) then the image of inertia is unipotent. Let \( \Sigma \) be the set of places outside 2 where \( \rho \) is ramified. If \( v \in \Sigma \) then

\[
\rho|_{G_{F'_v}} \cong \begin{pmatrix} \gamma_v \chi_{\text{cyc}} & \ast \\ 0 & \gamma_v \end{pmatrix},
\]

where \( \gamma_v \) is an unramified character such that \( \gamma_v^2 = \psi|_{G_{F'_v}} \).

Since \( \bar{\rho} \) is assumed to be modular, Lemma 3.29 implies that \( \bar{\rho} \cong \bar{\rho}_f \), where \( f \) is a Hilbert eigenform of parallel weight 2, and an unramified principal series at \( v \mid 2 \). Using Lemma 3.5.3 of [Kisin 2009c] (see also Theorem 8.4 of [Khare and Wintenberger 2009b]) there is an admissible base change \( F''/F' \) such that \( \rho|_{G_{F''}} \) is ramified at an even number of places outside 2. We still denote this set by \( \Sigma \), and there is a Hilbert eigenform \( g \) over \( F'' \) such that \( \bar{\rho}|_{G_{F''}} \cong \bar{\rho}_g \), and such that \( g \) has parallel weight 2, is special of conductor 1 at \( v \in \Sigma \), and is unramified otherwise.

Let \( D \) be the quaternion algebra with center \( F'' \) ramified exactly at all infinite places and all \( v \in \Sigma \). Choose a place \( v_1 \) of \( F'' \) as in Lemma 3.2 and such that \( \bar{\rho} \) is unramified at \( v_1 \) and \( \bar{\rho}((\text{Frob}_{v_1})) \) has distinct eigenvalues. Let \( S \) be the union of infinite places, \( \Sigma \), places above 2 and \( v_1 \). Let \( U = \prod v U_v \) be an open subgroup of \( (D \otimes_{F''} \mathbb{A}_{F''}^\times)^\times \) such that \( U_v = \mathcal{O}_D^\times \) if \( v \neq v_1 \) and \( U_{v_1} \) is unipotent upper triangular modulo \( \sigma_{v_1} \). We note that Lemma 3.2 implies that \( U \) satisfies (29). Let \( m \) be the maximal ideal in the Hecke algebra \( \mathcal{T}_{S,\mathcal{O}}^{\text{univ}} \) corresponding to \( \bar{\rho} \).

Let \( g^D \) be the eigenform on \( D \) corresponding to \( g \) via the Jacquet–Langlands correspondence. Then \( g^D \in S_{\sigma,\psi',(U,\mathcal{O})_m} \), where \( \sigma \) is the trivial representation of \( U \) and \( \psi'(\mathbb{A}_{F''}^\times) \to \mathcal{O}_D^\times \) is a suitable character congruent to \( \psi \) modulo \( \sigma \). In particular, \( S_{\sigma,\psi',(U,\mathcal{O})_m} \neq 0 \). It follows from Lemma 3.20 that \( S_{\sigma,\psi',(U,\mathcal{O})_m} \neq 0 \) for all \( \sigma = \bigotimes v|2 \sigma_v \), where \( \sigma_v \) is either \( \tilde{1} \) or \( \bar{1} \). Since \( U \) satisfies (29), we deduce that \( S_{\sigma,\psi'(U,\mathcal{O})_m} \neq 0 \) for any irreducible smooth \( k \)-representation \( \sigma \) of \( \prod v|2 \text{GL}_2(\mathbb{Z}_2) \). Since \( U \) satisfies (29), we deduce via Lemma 3.1.4 of [Kisin 2009c] that \( S_{\sigma,\psi'(U,\mathcal{O})_m} \neq 0 \) for any continuous finite-dimensional representation \( \sigma \) of \( \prod v|2 \text{GL}_2(\mathbb{Z}_2) \) on which \( U \cap (\mathbb{A}_{F''}^\times)^\times \) acts by \( \psi \).
For $v | 2$ suppose that $\rho|_{G_{F_v'}}$ has Hodge–Tate weights $w_v = (a_v, b_v)$ with $b_v > a_v$ and inertial type $\tau_v$. Let $\sigma_v$ be defined by (30) and let $\sigma = \otimes_{v|2} \sigma_v$. The above implies that $S_{\sigma, \psi}(U, \mathcal{O})_m \neq 0$ and, since $\rho|_{G_{F_v'}}$ defines a maximal ideal of $R_{F_v'}^\psi, S[{1 \over 2}]$, the assertion follows from Corollary 3.27.

Acknowledgements

The local part was written at about the same time as [Colmez et al. 2014], and it would not have happened if not for the competitive nature of the correspondence I had with Gabriel Dospinescu at the time. I thank Gaëtan Chenevier for the correspondence on 2-dimensional 2-adic determinants. The global part owes a great deal to the work of Khare and Wintenberger [2009b] and Kisin [2009a; 2009b; 2009c] as will be obvious to the reader. I thank Mark Kisin and Jean-Pierre Wintenberger for answering my questions about their work. I would like to especially thank Toby Gee for his explanations of the Taylor–Wiles–Kisin patching method and for pointing out the right places in the literature to me. I thank Lennart Gehrmann, Jochen Heinloth and Shu Sasaki for a number of stimulating discussions. I thank Gabriel Dospinescu, Matthew Emerton, Toby Gee and Jack Thorne for their comments on the earlier draft. I thank Patrick Allen for pointing out an error in the earlier draft, and for subsequent correspondence, which led to a fix. I thank the referees for their careful reading of the paper.

References


Communicated by Brian Conrad
Received 2015-09-01 Revised 2016-04-22 Accepted 2016-05-22

paskunas@uni-due.de

Universität Duisburg-Essen, Fakultät für Mathematik, Thea-Leymann-Str. 9, 45127 Essen, Germany
A probabilistic Tits alternative and probabilistic identities

Michael Larsen and Aner Shalev

We introduce the notion of a probabilistic identity of a residually finite group $\Gamma$. By this we mean a nontrivial word $w$ such that the probabilities that $w = 1$ in the finite quotients of $\Gamma$ are bounded away from zero.

We prove that a finitely generated linear group satisfies a probabilistic identity if and only if it is virtually solvable.

A main application of this result is a probabilistic variant of the Tits alternative: Let $\Gamma$ be a finitely generated linear group over any field and let $G$ be its profinite completion. Then either $\Gamma$ is virtually solvable, or, for any $n \geq 1$, $n$ random elements $g_1, \ldots, g_n$ of $G$ freely generate a free (abstract) subgroup of $G$ with probability 1.

We also prove other related results and discuss open problems and applications.

1. Introduction

The celebrated Tits alternative [1972] asserts that a finitely generated linear group is either virtually solvable or has a (nonabelian) free subgroup. A number of variations and extensions of this result have been obtained over the years. In particular, it is shown in [Breuillard and Gelander 2007] that if $\Gamma$ is a finitely generated linear group which is not virtually solvable then its profinite completion $\hat{\Gamma}$ has a dense free subgroup of finite rank (this answers a question from [Dixon et al. 2003], where a somewhat weaker result was obtained). The purpose of this paper is to establish a probabilistic version of the Tits alternative, and to relate it to the notion of probabilistic identities, which is interesting in its own right.

In order to formulate our first result, let us say that a profinite group $G$ is randomly free if for any positive integer $n$ the set of $n$-tuples in $G^n$ which freely generate a free subgroup of $G$ (isomorphic to $F_n$) has measure 1 (with respect to the normalized

---

Larsen was partially supported by NSF grant DMS-1401419. Shalev was partially supported by ERC advanced grant 247034, ISF grant 1117/13 and the Vinik Chair of Mathematics, which he holds.

**MSC2010:** primary 20G15; secondary 20E18.

**Keywords:** Tits alternative, residually finite, virtually solvable, probabilistic identity, profinite completion.
Haar measure on $G^n$). We also say that a (discrete) residually finite group $\Gamma$ is randomly free if its profinite completion is randomly free.

Recall that related notions have already been studied in various contexts. For example, Epstein [1971] showed that connected finite-dimensional nonsolvable real Lie groups are randomly free (in the sense that the set of $n$-tuples which do not freely generate a free subgroup has measure zero). Later it was shown by Szegedy [2005] that the Nottingham pro-$p$ group is randomly free (answering a question of the second author). Furthermore, Abért proved [2005] that some other groups are randomly free; these include the Grigorchuk group and profinite weakly branch groups.

We can now state our probabilistic Tits alternative.

**Theorem 1.1.** Let $\Gamma$ be a finitely generated linear group over any field. Then either $\Gamma$ is virtually solvable or $\Gamma$ is randomly free.

The proof of this result relies on the notion and properties of probabilistic identities which we introduce below.

Let $w = w(x_1, \ldots, x_n)$ be a nontrivial element of the free group $F_n$, and let $\Gamma$ be a residually finite group. Consider the induced word map $\Gamma^n \to \Gamma$, which, by a slight abuse of notation, we also denote $w$. If the image $w(\Gamma^n)$ of this map is $\{1\}$ then $w$ is an identity of $\Gamma$. We say that $w$ is a probabilistic identity of $\Gamma$ if there exists $\epsilon > 0$ such that, for each finite quotient $H = \Gamma/\Delta$ of $\Gamma$, the probability $P_H(w)$ that $w(h_1, \ldots, h_n) = 1$ (where the $h_i \in H$ are chosen independently with respect to the uniform distribution on $H$) is at least $\epsilon$. This amounts to saying that, in the profinite completion $G = \hat{\Gamma}$ of $\Gamma$, the probability (with respect to the Haar measure) that $w(g_1, \ldots, g_n) = 1$ is positive.

For example, $w = x_1^2$ is a probabilistic identity of the infinite dihedral group $\Gamma = D_{\infty}$, since in any finite quotient $\Gamma/\Delta = D_n$ of $\Gamma$ we have $P_{\Gamma/\Delta}(w) \geq \frac{1}{2}$. Note that, in this example, $w$ is not an identity on a finite index subgroup of $\Gamma$, but it is an identity on a coset of the cyclic subgroup of index two.

More generally, probabilistic identities may be regarded as an extension of the notion of coset identities. Recall that a word $1 \neq w \in F_n$ is said to be a coset identity of the infinite group $\Gamma$ if there exists a finite index subgroup $\Delta \leq \Gamma$ and cosets $\gamma_1 \Delta, \ldots, \gamma_n \Delta$ (where $\gamma_i \in \Gamma$) such that $w(\gamma_1 \Delta, \ldots, \gamma_n \Delta) = \{1\}$.

Our main result on probabilistic identities is the following.

**Theorem 1.2.** A finitely generated linear group satisfies a probabilistic identity if and only if it is virtually solvable.

Theorem 1.2 has several consequences. First, it easily implies Theorem 1.1. To show this, suppose $\Gamma$ is not virtually solvable, and let $G$ be the profinite completion of $\Gamma$. Note that $g_1, \ldots, g_n \in G$ freely generate a free subgroup of $G$ if and only if $w(g_1, \ldots, g_n) \neq 1$ for every $1 \neq w \in F_n$. By Theorem 1.2 above, the probability
that \( w(g_1, \ldots, g_n) = 1 \) is 0 for any such \( w \). As Haar measure is \( \sigma \)-additive, the probability that there exists \( w \neq 1 \) such that \( w(g_1, \ldots, g_n) = 1 \) is also 0. Thus, \( g_1, \ldots, g_n \) freely generate a free subgroup with probability 1, proving Theorem 1.1.

Secondly, Theorem 1.2 immediately implies the following.

**Corollary 1.3.** A finitely generated linear group which satisfies a probabilistic identity satisfies an identity.

It would be interesting to find out whether the same holds without the linearity assumption. We discuss this and related problems and applications in Section 3.

In the course of the proof of Theorem 1.2 we establish a result of independent interest, showing that probabilistic identities on finitely generated linear groups are in fact coset identities.

The arguments proving this result also prove a more general result on probabilistic identities with parameters. Let \( w(x_1, \ldots, x_n, y_1, \ldots, y_m) \) be a word in the variables \( x_1, \ldots, x_n, y_1, \ldots, y_m \), and let \( \gamma_1, \ldots, \gamma_m \) be elements of a residually finite group \( \Gamma \). Consider the word with parameters \( v(x_1, \ldots, x_n) := w(x_1, \ldots, x_n, \gamma_1, \ldots, \gamma_m) \).

The notions of a probabilistic identity with parameters and of a coset identity with parameters are then defined in the obvious way.

Note that Theorem 1.2 cannot be generalized to probabilistic identities with parameters. For example, let \( \gamma_1 \in \Gamma \) be a central element. Then the word with parameters \([x_1, \gamma_1]\) is an identity on \( \Gamma \), though \( \Gamma \) need not be virtually solvable. However, we can show the following.

**Theorem 1.4.** Let \( \Gamma \) be a finitely generated linear group over any field. Then every probabilistic identity (possibly with parameters) on \( \Gamma \) is a coset identity.

It easily follows that, if \( w \) is a word in \( n \) variables (possibly with parameters from \( \Gamma \)), and \( \gamma \in \Gamma \) is such that in all finite quotients \( H = \Gamma / \Delta \) of \( \Gamma \) the probability that \( w(h_1, \ldots, h_n) = \gamma + \Delta \) is at least some fixed \( \epsilon > 0 \), then the fiber \( w^{-1}(\gamma) \) contains the Cartesian product \( \gamma_1 \Delta \times \cdots \times \gamma_n \Delta \) of cosets of some finite index subgroup \( \Delta \leq \Gamma \). Indeed, apply Theorem 1.4 to the word with parameters \( w_{\gamma}^{-1} \).

In fact, the proof of Theorem 1.4 gives rise to an even more general result of independent interest. In order to formulate it, let \( \Gamma \) be a linear group and let \( n \) be a positive integer. Let us say that a subset \( \Xi \) of \( \Gamma^n \) is Zariski-closed if there is an embedding of \( \Gamma \) in \( \text{GL}_r(F) \) (for some field \( F \) and a positive integer \( r \)) and a Zariski-closed subset \( X \) of \( \text{GL}_r^n \) such that \( \Xi = X(F) \cap \Gamma^n \).

Then we have the following.

**Theorem 1.5.** Let \( \Gamma \) be a finitely generated linear group over any field, and let \( n \geq 1 \). Let \( \Xi \subseteq \Gamma^n \) be a Zariski-closed subset. Suppose there exists \( \epsilon > 0 \) such that \( |\Xi \Delta^n / \Delta^n| \geq \epsilon |\Gamma / \Delta|^n \) for all normal subgroups of finite index \( \Delta \) of \( \Gamma \). Then there exists a finite index subgroup \( \Delta \leq \Gamma \) and elements \( \gamma_1, \ldots, \gamma_n \in \Gamma \) such that \( \Xi \supseteq \gamma_1 \Delta \times \cdots \times \gamma_n \Delta \).
This result is proved using an easy adaptation of the proof of Theorem 1.4, which we leave for the interested reader. Theorem 1.5 amounts to saying that if the closure of $\Xi$ in the profinite group $(\hat{\Gamma})^n$ has positive Haar measure, then it has a nonempty interior.

It is shown in [Breuillard and Gelander 2007, Theorem 8.4] that a finitely generated linear group which satisfies a coset identity (without parameters) is virtually solvable. Using this result we can immediately deduce Theorem 1.2 from Theorem 1.4. In fact, we provide here a self-contained proof of Theorem 1.2 using Theorem 1.4 and Proposition 2.5 below.

Our original approach to proving Theorem 1.2 relied on strong approximation for linear groups and on establishing upper bounds on the probabilities $P_G(w)$, where $G$ is a group satisfying $T^k \leq G \leq \text{Aut}(T^k)$ for a finite simple group $T$. However, this approach is rather involved. A shorter and simpler proof of Theorems 1.4 and 1.2 is given in Section 2.

The idea is to use linearity to map $\Gamma$ into a “linear algebraic group” $G$ over an infinite product $\prod_m A/m$ of finite fields. The closure of the image is then a profinite group. Suppose that for some Zariski-closed subset $X \subset G^n$, the measure of the closure of $X (\prod_m A/m) \cap \Gamma^n$ is positive. Every translate of $X$ by an element of $\Gamma^n$ has the same property. Unless $X$ is a union of connected components of $G^n$ we can find an infinite set of pairwise distinct translates of $X$, each of which has the same positive-measure property. Thus, some pairs of translates of $X$ must intersect $\Gamma^n$ with positive measure; intersecting $X$ with a suitable translate by an element of $\Gamma^n$, we obtain a proper closed subset of $X$ with the same property as $X$ itself. This process cannot continue indefinitely. The theorem is obtained by applying it to the fiber over 1 of a nontrivial word map $w$. The actual implementation uses the language of (affine) schemes and a notion somewhat weaker than that of measure.

In fact, this method of proof, and Proposition 2.5 in particular, yields the following extension of Theorem 1.2: Suppose $\Gamma$ is a finitely generated linear group which is not virtually solvable. Then all fibers in $(\hat{\Gamma})^n$ of all nontrivial words $w \in F_n$ have measure 0.

In other words, for a finite group $H$, let $P_{H,w}$ denote the probability distribution induced on $H$ by $w$ (so that, for $h \in H$, $P_{w,H}(h)$ is the probability that $w(h_1, \ldots, h_n) = h$). Its $\ell_\infty$-norm is defined by $\|P_{H,w}\|_\infty = \max_{h \in H} P_{H,w}(h)$. Then we have:

**Theorem 1.6.** Let $\Gamma$ be a finitely generated linear group. Suppose for some $n \geq 1$ and $1 \neq w \in F_n$ there exists $\epsilon > 0$ such that for all finite quotients $H$ of $\Gamma$ we have $\|P_{H,w}\|_\infty \geq \epsilon$. Then $\Gamma$ is virtually solvable.

See also [Aoun 2011] for a different probabilistic Tits alternative, related to certain random walks on the discrete linear group $\Gamma$. 
2. Proof of Theorems 1.4 and 1.2

If a group \( \Gamma \) acts on a topological space \( X \) and \( Y \subseteq X \), we say \( Y \) is \( \Gamma \)-finite if its orbit under \( \Gamma \) is finite. We say a closed subset \( Z \subseteq X \) is \( \Gamma \)-covered by \( Y \) if \( Z \) is a closed subset of some finite union of \( \Gamma \)-translates of \( Y \).

**Lemma 2.1.** Let \( \Gamma \) be a group acting on a set \( X \). If \( Y_1, \ldots, Y_n \) are subsets of \( X \) which are not \( \Gamma \)-finite, then there exists \( g \in \Gamma \) such that \( gY_i \neq Y_j \) for \( 1 \leq i, j \leq n \).

**Proof.** For given \( i, j \), the set of \( g \) such that \( gY_i = Y_j \) is either empty or is a left coset of the stabilizer of \( Y_i \) in \( \Gamma \). By a theorem of B. H. Neumann [1954], a group cannot be covered by a finite collection of left cosets of subgroups of infinite index. The result follows. \( \square \)

**Proposition 2.2.** Let \( X \) be a Noetherian topological space and \( \Gamma \) a group of homeomorphisms \( X \to X \). Let \( f \) denote a function from the set of closed subsets of \( X \) to \([0, 1]\) satisfying the following conditions:

(I) If \( Z \subseteq Y \) are closed subsets of \( X \), then \( f(Z) \leq f(Y) \).

(II) For all closed subsets \( Y \subseteq X \) and all \( g \in \Gamma \) such that \( f(Y \cap gY) = 0 \), we have \( f(Y \cup gY) \geq 2f(Y) \).

If \( Y \subseteq X \) is closed and \( \Gamma \)-covers some closed subset \( W \subseteq X \) with \( f(W) > 0 \), then \( Y \) \( \Gamma \)-covers some closed \( \Gamma \)-stable subset \( Z \subseteq X \) with \( f(Z) > 0 \).

**Proof.** By the Noetherian hypothesis, we may assume without loss of generality that \( Y \) is minimal for the property of \( \Gamma \)-covering a set of positive \( f \)-value. If two distinct irreducible components \( Y_i \) and \( Y_j \) of \( Y \) were \( \Gamma \)-translates of one another, we could replace \( Y \) with the union of all of its components except \( Y_j \), and the resulting closed set would still \( \Gamma \)-cover a set of positive \( f \)-value. This is impossible by the minimality of \( Y \).

If \( Y \) is \( \Gamma \)-finite, then

\[
Z := \bigcup_{g \in \Gamma} gY
\]

is a \( \Gamma \)-stable finite union of \( \Gamma \)-translates of \( Y \) containing \( W \). By condition (I), it satisfies \( f(Z) > 0 \), so we are done. As \( Y \) is a finite union of irreducible components, we may therefore assume at least one such component \( Y_0 \) is not \( \Gamma \)-finite. We write \( Y = Y_0 \cup Y' \), where no \( \Gamma \)-translate of \( Y' \) contains \( Y_0 \).

By condition (I), there exists a finite sequence \( g_1, \ldots, g_r \in \Gamma \) such that \( f(Z) > 0 \) for

\[
Z := g_1Y \cup \cdots \cup g_rY.
\]
We choose the \( g_i \) so that
\[
    f(Z) > \frac{\sup_{\Delta \subseteq \Gamma \text{ finite}} f(\bigcup_{g \in \Delta} gY)}{2}. \tag{2-1}
\]

As no \( \Gamma \)-translate of \( Y_0 \) is \( \Gamma \)-finite, Lemma 2.1 implies that there exists \( g \) such that \( g_i Y_0 \neq gg_j Y_0 \) for all \( i, j \). Thus,
\[
    Y' \cup \bigcup_{i,j} (Y_0 \cap g_i^{-1} gg_j Y_0) \subset Y
\]
\( \Gamma \)-covers \( Z \cap gZ \). By the minimality of \( Y \), this means \( f(Z \cap gZ) = 0 \). By condition (II), \( f(Z \cup gZ) \geq 2f(Z) \), which contradicts (2-1). We conclude that \( Z \) must be \( \Gamma \)-finite. \( \square \)

Now, let \( A \) be an integral domain finitely generated over \( \mathbb{Z} \) with fraction field \( K \). Let \( \mathcal{G} = \text{Spec } B \) be an affine group scheme of finite type over \( A \) (see [Waterhouse 1979]). As usual, for every commutative \( A \)-algebra \( T \), let \( \mathcal{G}(T) \) denote the set of \( \text{Spec } T \)-points of \( \mathcal{G} \to \text{Spec } A \), i.e., the set of \( A \)-algebra homomorphisms \( B \to T \). The group structure on \( \mathcal{G} \) makes each \( \mathcal{G}(T) \) a group, functorially in \( T \). We regard \( \mathcal{G} \) as a topological space with respect to its Zariski topology. If \( Y \subseteq \mathcal{G} \) is a closed subset, we define \( Y(T) \) to be the subset of \( \mathcal{G}(T) \) consisting of \( A \)-homomorphisms \( B \to T \) such that the corresponding map of topological spaces \( \text{Spec } T \to \mathcal{G} \) sends \( \text{Spec } T \) into a subset of \( Y \). If \( Z \subseteq \mathcal{G} \) is another closed subset, then
\[
(Y \cap Z)(T) = Y(T) \cap Z(T),
\]
but, in general, the inclusion
\[
Y(T) \cup Z(T) \subseteq (Y \cup Z)(T)
\]
need not be an equality.

We define
\[
P(\mathcal{G}, A) := \prod_{m \in \text{Maxspec}(A)} \mathcal{G}(A/m),
\]
where Maxspec denotes the set of maximal ideals, and \( P(\mathcal{G}, A) \) is endowed with the product topology. Note that as \( \mathcal{G} \) is of finite type (i.e., \( B \) is a finitely generated \( A \)-algebra) and every \( A/m \) is a field finitely generated over \( \mathbb{Z} \) (and hence finite), it follows that each \( \mathcal{G}(A/m) \) is finite and \( P(\mathcal{G}, A) \) is a profinite group. For any closed subset \( X \subseteq \mathcal{G} \), we define the closed subset
\[
P(X, A) := \prod_{m \in \text{Maxspec}(A)} X(A/m) \subseteq P(\mathcal{G}, A).
\]

**Lemma 2.3.** If \( X \subseteq \mathcal{G} \) does not meet the generic fiber \( \text{Spec } B \otimes_A K \subset \mathcal{G} \), then \( P(X, A) \) is empty.
Proof. If $I \subseteq B$ is the ideal defining $X$, then $(B/I) \otimes_A K = 0$, so $I \otimes_A K = B \otimes_A K$. It follows that there exist elements $b_i \in I$ and $a_i/a'_i \in K$ such that
\[
\sum_i b_i \otimes a_i/a'_i = 1,
\]
and clearing denominators we see that some nonzero element $a' := \prod_i a'_i \in A$ belongs to $I$. If $m$ is a maximal ideal of $A[1/a']$, then $A[1/a']/m$ is a field finitely generated over $\mathbb{Z}$, hence a finite field, and therefore $m \cap A$ is a maximal ideal of $A$. Thus, the image of $a'$ in $A/(m \cap A)$ is nonzero, from which it follows that there are no $A$-homomorphisms $B/I \to A/(m \cap A)$, i.e., $X(A/(m \cap A)) = \emptyset$. □

For any subgroup $\Gamma \subseteq G(A) \subseteq P(G, A)$, we define $\overline{\Gamma}$ to be the closure of $\Gamma$ in $P(G, A)$. This is a closed subgroup of a profinite group and therefore a profinite group itself. We endow it with Haar measure $\mu_\Gamma$, normalized so that $(\overline{\Gamma}, \mu_\Gamma)$ is a probability space. In particular, left translation by $\Gamma$ is a continuous measure-preserving action on $(\overline{\Gamma}, \mu_\Gamma)$. As Haar measure is outer regular, for every Borel set $B$,
\[
\mu_\Gamma(B) = \inf_{S \subseteq \text{Maxspec}(A)} \frac{|\text{pr}_S B|}{|\text{pr}_S \Gamma|},
\]
where $S$ ranges over all finite sets of maximal ideals of $A$ and $\text{pr}_S$ denotes projection onto $\prod_{m \in S} G(A/m)$.

For any positive integer $n$, we let $G^n$ denote the $n$-th fiber power of $G$ relative to $A$, i.e., defining
\[
B_n := B \otimes_A B \otimes_A \cdots \otimes_A B,
\]
we define $G^n := \text{Spec } B_n$, regarded as a topological space with respect to the Zariski topology. Note that in general the Zariski topology on $G^n$ is not the product topology. However, by the universal property of tensor products, $G^n(T)$ is canonically isomorphic to $G(T)^n$ for all commutative $A$-algebras $T$. Moreover, $B_n$ is a finitely generated $\mathbb{Z}$-algebra, and by the Hilbert basis theorem this implies that $G^n$ is a Noetherian topological space.

We consider the closure $\overline{\Gamma^n}$ of $\Gamma^n$ in $P(G^n, A)$. For any closed subset $Y \subseteq G^n$, we define
\[
P_\Gamma(Y) := \overline{\Gamma^n} \cap P(Y, A).
\]
Thus, if $Y$ and $Z$ are closed subsets of $G^n$,
\[
P_\Gamma(Y \cap Z) = \overline{\Gamma^n} \cap P(Y \cap Z, A) = \overline{\Gamma^n} \cap (P(Y, A) \cap P(Z, A)) = P_\Gamma(Y) \cap P_\Gamma(Z).
\]
As
\[
P(Y \cup Z, A) = \prod_{m \in \text{Maxspec}(A)} (Y(A/m) \cup Z(A/m)) \supseteq P(Y, A) \cup P(Z, A),
\]
we have
\[
P_\Gamma(Y \cup Z) \supseteq P_\Gamma(Y) \cup P_\Gamma(Z).
\]
Defining
\[
f(Y) := \mu_\Gamma_n(P_\Gamma(Y)),
\]
condition (I) of Proposition 2.2 is obvious. As \(\mu_\Gamma_n\) is a measure, if \(f(Y \cap Z) = 0\), then
\[
f(Y \cup Z) = \mu_\Gamma_n(P_\Gamma(Y \cup Z))
\]
\[
\geq \mu_\Gamma_n(P_\Gamma(Y) \cup P_\Gamma(Z))
\]
\[
= \mu_\Gamma_n(P_\Gamma(Y)) + \mu_\Gamma_n(P_\Gamma(Z)) - \mu_\Gamma_n(P_\Gamma(Y) \cap P_\Gamma(Z))
\]
\[
= f(Y) + f(Z) - f(Y \cap Z) = f(Y) + f(Z).
\]
As \(\mu_\Gamma_n\) is \(\Gamma^n\)-invariant, this implies condition (II).

**Proposition 2.4.** Let \(G\) denote a linear algebraic group over a field \(K\). If \(\Gamma\) is Zariski-dense in \(G(K)\), then a nonempty closed subset \(Y\) of \(G^n\) is \(\Gamma^n\)-finite if and only if it is a union of connected components of \(G^n\).

**Proof.** If \(Y\) is \(\Gamma^n\)-finite, its stabilizer \(\Delta\) is of finite index in \(\Gamma^n\), which implies that the Zariski closure \(D\) of \(\Delta\) in \(G^n\) has finite index in \(G^n\). Thus \(D \cap (G^n)^\circ\) is of finite index in \((G^n)^\circ\). As \((G^n)^\circ\) is connected, it follows that \(D\) contains \((G^n)^\circ\). The Zariski closure of any left coset of \(\Gamma^n\) is a left coset of \(D\) and therefore a union of cosets of \((G^n)^\circ\). Conversely, any left translate of a coset of \((G^n)^\circ\) is again such a coset, so the orbit of any union of connected components of \(G^n\) is finite. \(\Box\)

We can now prove Theorem 1.4.

**Proof.** We fix a faithful representation \(\rho : \Gamma \to \text{GL}_r(F)\), where \(F\) is an algebraically closed field. Let \(G \subseteq \text{GL}_r\) denote the Zariski closure of \(\Gamma\) in \(\text{GL}_r\).

We recall how to extend \(G\) to a subgroup scheme of \(\text{GL}_r\) defined over a finitely generated \(\mathbb{Z}\)-algebra. Let
\[
R_{\mathbb{Z}, r} := \mathbb{Z}[x_{ij}, y]_{i, j = 1, \ldots, r}/(y \det(x_{ij}) - 1)
\]
denote the coordinate ring of \(\text{GL}_r\) over \(\mathbb{Z}\), and let
\[
\Delta_{\mathbb{Z}, r} : R_{\mathbb{Z}, r} \to R_{\mathbb{Z}, r} \otimes_{\mathbb{Z}} R_{\mathbb{Z}, r}, \quad S_{\mathbb{Z}, r} : R_{\mathbb{Z}, r} \to R_{\mathbb{Z}, r}, \quad \text{and} \quad \epsilon_{\mathbb{Z}, r} : R_{\mathbb{Z}, r} \to \mathbb{Z}
\]
denote the ring homomorphisms associated to the multiplication, inverse, and unit maps. Closed subschemes of \(\text{GL}_r\) over any commutative ring \(A\) are in one-to-one
correspondence with ideals \( I \) of \( R_{A,r} := A \otimes \mathbb{Z} R_{\mathbb{Z},r} \), and such an ideal defines a group subscheme if and only if \( I \) is a Hopf ideal [Waterhouse 1979, §2.1], i.e., if and only if it satisfies the following three conditions:

\[
\Delta_{A,r}(I) \subseteq I \otimes_A R_{A,r} + R_{A,r} \otimes_A I,
\]

\[
S_{A,r}(I) \subseteq I,
\]

\[
\epsilon_{A,r}(I) = \{0\}.
\]

We fix a finite set of generators \( h_k \) of the ideal \( I_F \) in \( R_{F,r} \) associated to \( G \) as a closed subvariety of \( \text{GL}_r \) over \( F \). We lift each \( h_k \) to an element \( \tilde{h}_k \in F[x_{ij}, y] \). For any subring \( A \subseteq F \) such that \( \tilde{h}_k \in A[x_{ij}, y] \), we denote again by \( h_k \) the image of \( \tilde{h}_k \) in \( R_{A,r} \); this should not cause confusion. Let \( A_0 \) denote the subring of \( F \) generated by all matrix entries in \( \text{GL}_r(F) \) of the \( \rho(g_j) \), as \( g_j \) runs over some finite generating set of \( \Gamma \), together with all coefficients of the \( \tilde{h}_k \). Let \( I_0 \) denote the ideal generated by the elements \( h_k \) in \( R_{A_0,r} \), and let \( K \) denote the fraction field of \( A_0 \). As

\[
\Delta_{A_0,r}(I_0) \subseteq I_0 \otimes_{A_0} R_{K,r} + R_{K,r} \otimes_{A_0} I_0
\]

and

\[
S_{A_0,r}(I_0) \subseteq I_0 \otimes_{A_0} R_{K,r},
\]

there exists \( a \in A_0 \) such that

\[
\Delta_{A_0,r}(h_i) \in I_0 \otimes_{A_0} R_{A_0[1/a],r} + R_{A_0[1/a],r} \otimes_{A_0} I_0
\]

and

\[
S_{A_0,r}(h_i) \in I_0 \otimes_{A_0} A_0[1/a]
\]

for all \( i \), and therefore, setting \( A := A_0[1/a] \) and \( I := I_0 \otimes_{A_0} A \), we have that \( I \) is a Hopf ideal of \( R_{A,r} \). We set \( \mathcal{G} := \text{Spec} R_{A,r}/I \), the closed group subscheme of \( \text{GL}_r \) over \( A \) defined by \( h_k \in R_{A,r} \). By construction, \( \rho(\Gamma) \) is a Zariski-dense finitely generated subgroup of \( \mathcal{G}(A) \).

Now, let \( w \) be a probabilistic identity on \( \Gamma \) (possibly with parameters). Consider \( w \) as a morphism of schemes over \( A \) from \( \mathcal{G}^n \) to \( \mathcal{G} \). Let \( Y := w^{-1}(1) \subseteq \mathcal{G}^n \). We define \( f \) as above. If \( f(Y) > 0 \), then \( Y \) \( \Gamma \)-covers a set of positive \( f \)-value, so by Proposition 2.2 \( Y \) \( \Gamma \)-covers a closed \( \Gamma \)-stable subset \( Z \) with \( f(Z) > 0 \). By Lemma 2.3, \( Z \) must meet the generic fiber \( G^n \) of \( \mathcal{G}^n \), which implies that \( Y \) must meet the generic fiber. Proposition 2.4 now implies that \( Z \cap G^n \) contains a connected component of \( G^n \), and it follows that \( Y \cap G^n \) contains a connected component, i.e., \( w \) is a coset identity. Thus, we may assume \( f(Y) = 0 \).

Therefore, for every \( \epsilon > 0 \), there exists a finite set \( S \) of maximal ideals of \( A \) such that

\[
\frac{|\text{pr}_S w^{-1}(1)|}{|\text{pr}_S \Gamma^n|} < \epsilon.
\]
Defining $\Delta$ to be the kernel of $\text{pr}_S$, we see that, in the finite quotient $\Gamma/\Delta$, the probability that the word map $w$ attains the value $1 + \Delta$ is less than $\epsilon$. It follows that $w$ is not a probabilistic identity on $\Gamma$. This contradiction completes the proof of Theorem 1.4.

**Proposition 2.5.** Let $K$ be a field and $G$ a linear algebraic group over $K$ with nontrivial adjoint semisimple identity component. Let $w \in F_n$ be a nontrivial word and let $g_0 \in G(K)$. Then $w^{-1}(g_0)$ does not contain any connected component of $G^n$.

**Proof.** Equivalently, we claim that $\dim w^{-1}(g_0) < \dim G^n$. Since dimensions do not depend on the base field, we may and shall assume, without loss of generality, that $K$ is algebraically closed. Let $G^\circ$ be the identity component, $T$ a maximal torus of $G^\circ$ and $B$ a Borel subgroup of $G^\circ$ containing $T$. Let $\Phi$ be the root system of $G$ with respect to $T$, and let $\Phi^+$ denote the set of roots of $B$ with respect to $T$. Every maximal torus of $G^\circ$ is conjugate under $G^\circ(K)$ to $T$. The Weyl group $N_G(T)/T$ acts transitively on the set of Weyl chambers, so every pair $T' \subset B'$ is conjugate to $T \subset B$ by some element of $G^\circ(K)$. In particular, for any $g \in G(K)$, the pair $g^{-1}Tg \subset g^{-1}Bg$ is conjugate in $G^\circ(K)$ to $T \subset B$, or, equivalently, there is some element $h \in gG^\circ(K)$ such that conjugation by $h$ stabilizes $T$ and $B$. The highest root $\alpha$ of $\Phi^+$ is determined by $B$, so $h$ likewise preserves $\alpha$. It therefore normalizes $\ker \alpha^\circ$, and therefore the derived group $G_\alpha$ of the centralizer of $\ker \alpha^\circ$. This group is semisimple and of type $A_1$, so every element that normalizes it acts by an inner automorphism. It follows that the centralizer of $G_\alpha$ in $G$ meets every connected component of $G$.

Suppose now that $w$ is constant on $g_1 G^\circ \times \cdots \times g_n G^\circ$ for some $g_1, \ldots, g_n \in G(K)$. Without loss of generality we may assume that all $g_i$ centralize $G_\alpha$. As $w$ is constant on $g_1 G_\alpha \times \cdots \times g_n G_\alpha$, and as

$$w(g_1 h_1, \ldots, g_n h_n) = w(g_1, \ldots, g_n) w(h_1, \ldots, h_n)$$

for all $h_1, \ldots, h_n \in G_\alpha(K)$, it follows that $w$ is constant on $G_\alpha^n$. This is impossible because nontrivial words give nontrivial word maps on all semisimple algebraic groups [Borel 1983].

**Proof of Theorem 1.2.** Every virtually solvable linear group satisfies a nontrivial identity. In the other direction, if $\Gamma \subset \text{GL}_r(K)$ satisfies a probabilistic identity, then it satisfies a coset identity by Theorem 1.4, and the same is true for its Zariski closure $G$. If $R$ denotes the maximal solvable normal subgroup of $G^\circ$, then $G/R$ also satisfies a coset identity, and by Proposition 2.5 this implies that $G/R$ is finite, i.e., that $G$ is virtually solvable, and so is $\Gamma$.

**3. Open problems**

In this section we discuss related open problems concerning finite and residually finite groups.
**Problem 3.1.** Do all finitely generated residually finite groups which satisfy a probabilistic identity satisfy an identity?

We also pose a related, finitary version of Problem 3.1.

**Problem 3.2.** Is it true that, for any word \( w \neq v \in F_n \) (for some \( m \)) such that, if \( G \) is a finite \( d \)-generated group satisfying \( P_G(w) \geq \epsilon \), then \( v \) is an identity of \( G \)?

Clearly, a positive answer to Problem 3.2 implies a positive answer to Problem 3.1. Both seem to be very challenging questions, which might have negative answers in general. However, in some special cases they are solved affirmatively. For example, if \( w = [x_1, x_2] \) or \( w = x_1^2 \), then it is known (see [Neumann 1989] and [Mann 1994]) that, for a finite group \( G \), if \( P_G(w) \geq \epsilon > 0 \), then \( G \) is bounded-by-abelian-by-bounded (in terms of \( \epsilon \)). This implies affirmative answers to Problems 3.1 and 3.2 for these particular words \( w \).

In general we cannot answer these problems for words of the form \( x_1^k \) \((k > 2)\). However, for a prime \( p \), a result of Khukhro [1986] shows that, if \( G \) is a finitely generated pro-\( p \) group satisfying a coset identity \( x_1^p \) (namely, there is a coset of an open subgroup consisting of elements of order \( p \) or 1) then \( G \) is virtually nilpotent (and hence satisfies an identity).

Another positive indication is the result showing that for a (nonabelian) finite simple group \( T \) and a nontrivial word \( w \) we have \( P_T(w) \to 0 \) as \( |T| \to \infty \) (see [Dixon et al. 2003] for this result, and also [Larsen and Shalev 2012] for upper bounds on \( P_T(w) \) of the form \( |T|^{-\alpha w} \)). This implies that a finite simple group \( T \) satisfying \( P_T(w) \geq \epsilon > 0 \) is of bounded size, hence it satisfies an identity (depending on \( w \) and \( \epsilon \) only).

Affirmative answers to Problems 3.1 and 3.2 would have far reaching applications. The argument proving Theorem 1.1 above also proves the following.

**Proposition 3.3.** Assume Problem 3.1 has a positive answer, and let \( \Gamma \) be a finitely generated residually finite group. Then either \( \Gamma \) satisfies an identity or \( \Gamma \) is randomly free.

In particular:

(i) If \( \Gamma \) does not satisfy an identity then \( \hat{\Gamma} \) has a nonabelian free subgroup.

(ii) If \( \hat{\Gamma} \) has a nonabelian free subgroup then almost all \( n \)-tuples in \( \hat{\Gamma} \) freely generate a free subgroup.

The next application concerns residual properties of free groups. It is well known that the free group \( F_n \) is residually-\( p \). But when is it residually \( X \) for a collection \( X \) of finite \( p \)-groups? If this is the case, then \( F_n \) is also residually \( Y \), where \( Y \) is the subset of \( X \) consisting of \( n \)-generated \( p \)-groups. Thus we may replace \( X \) by \( Y \) and assume all \( p \)-groups in \( X \) are \( n \)-generated. It is also clear that if \( F_n \) \((n > 1)\) is
residually $X$ then the groups in $X$ do not satisfy a common identity (namely, they generate the variety of all groups).

It turns out that, assuming an affirmative answer to Problem 3.2, these conditions are also sufficient.

**Proposition 3.4.** Assume Problem 3.2 has a positive answer. Let $n \geq 2$ be an integer, $p$ a prime, and $X$ a set of $n$-generated finite $p$-groups. Then the free group $F_n$ is residually $X$ if and only if the groups in $X$ do not satisfy a common identity.

To prove this, suppose the groups in $X$ do not satisfy a common identity. To show that $F_n$ is residually $X$, we have to find, for each $1 \neq w = w(x_1, \ldots, x_n) \in F_n$, a group $G \in X$ and an epimorphism $\phi : F_n \to G$, such that $\phi(w) \neq 1$. This amounts to finding a group $G \in X$ and an $n$-tuple $g_1, \ldots, g_n \in G$ generating $G$ such that $w(g_1, \ldots, g_n) \neq 1$ (and then $\phi$ is defined by sending $x_i$ to $g_i$). Suppose, given $w$, that there is no $G \in X$ with such an $n$-tuple. Then, for every $G \in X$, and every generating $n$-tuple $(g_1, \ldots, g_n) \in G^n$, we have $w(g_1, \ldots, g_n) = 1$. Now, the probability that a random $n$-tuple in $G^n$ generates $G$ is the probability that its image in $V^n$ spans $V$, where $V = G/\Phi(G)$ is the Frattini quotient of $G$, regarded as a vector space of dimension $\leq n$ over the field with $p$ elements. This probability is at least $\epsilon := \prod_{i=1}^n (1 - p^{-i}) > 0$. Thus $P_G(w) \geq \epsilon$ for all $G \in X$. By the affirmative answer to Problem 3.2, all the groups $G \in X$ satisfy a common identity $v \neq 1$ (which depends on $w$, $n$ and $p$). This contradiction proves Proposition 3.4.

This argument can be generalized to cases when $X$ consists of finite groups $G$ with the property that $n$ random elements of $G$ generate $G$ with probability bounded away from zero. See [Jaikin-Zapirain and Pyber 2011] and the references therein for the description of such groups and the related notion of positively finitely generated profinite groups.

**Acknowledgement**

We would like to acknowledge the referees’ very helpful suggestions, which led to some improvements in the paper.

**References**


A probabilistic Tits alternative and probabilistic identities


Communicated by Efim Zelmanov
Received 2015-10-29 Revised 2016-05-01 Accepted 2016-05-31

mjarsen@indiana.edu Department of Mathematics, Indiana University, Rawles Hall, Bloomington, IN 47405-5701, United States

shalev@math.huji.ac.il Einstein Institute of Mathematics, The Hebrew University of Jerusalem, 91904 Jerusalem, Israel
Authors may submit manuscripts in PDF format on-line at the Submission page at the ANT website.

**Originality.** Submission of a manuscript acknowledges that the manuscript is original 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 in ANT are usually in English, but articles written in other languages are welcome.

**Length** There is no a priori limit on the length of an ANT article, but ANT considers long articles only if the significance-to-length ratio is appropriate. Very long manuscripts might be more suitable elsewhere as a memoir instead of a journal article.

**Required items.** A brief abstract of about 150 words or less must be included. It should be self-contained and not make any reference to the bibliography. 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 subject classifications for the article, and, for each author, postal address, affiliation (if appropriate), and email address.

**Format.** Authors are encouraged to use LATEX but submissions in other varieties of TEX, and exceptionally in other formats, are acceptable. Initial uploads should 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 details about how your graphics were generated.

**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.
Modular elliptic curves over real abelian fields and the generalized Fermat equation

$\ell^2 + y^{2m} = z^n$

Samuele Anni and Samir Siksek

Geometry and stability of tautological bundles on Hilbert schemes of points

David Stapleton

Anabelian geometry and descent obstructions on moduli spaces

Stefan Patrikis, José Felipe Voloch and Yuri G. Zarhin

On the local Tamagawa number conjecture for Tate motives over tamely ramified fields

Jay Daigle and Matthias Flach

Heegner divisors in generalized Jacobians and traces of singular moduli

Jan Hendrik Bruinier and Yingkun Li

On 2-dimensional 2-adic Galois representations of local and global fields

Vytautas Paškūnas

A probabilistic Tits alternative and probabilistic identities

Michael Larsen and Aner Shalev